Letter to the Editor

Autosomal STR allele frequencies for the CODIS system from a large random population sample in Chile

Dear Editor,

The thirteen autosomal STR loci of the CODIS system were typed from DNA of 732 unrelated male individuals sampled from different locations in Chile. This is the first report of allele frequencies for the thirteen STRs loci defined in the CODIS system from the Chilean population.

Buccal swab samples were obtained from individuals requesting private paternity testing studies to TAAG Genetics laboratory between January 2008 and December 2009. To avoid any obvious bias in the population sampling, only alleged father samples from each paternity study were selected. Samples used in this study have a provenance from 54 different cities in Chile, from Arica to Punta Arenas, with a 52% of samples from Santiago. Each individual sampled in this study has signed his informed consent.

DNA extraction was performed from buccal swab samples with the DNA IQTM system (Promega, USA). The total amount of genomic DNA in each sample was quantified with the Nucleic dotMetricTM (G-Biosciences, USA) or the Quant-iTTM High Sensitivity DNA Assay kits (InvitrogenTM, USA). A total of 5 ng was used as template DNA for polymerase chain reaction (PCR) amplification. The 13 autosomal CODIS polymorphic STR loci (vWA, D7S820, CSF1PO, FGA, D18S51, D3S1358, D16S539, TPOX, D13S317, D8S1179, D21S11, TH01 and D5S818) and the AMEL sexual marker were amplified using the commercial M14-Multi-TAAG™ multiplex PCR amplification kit according to manufacturer instructions (TAAG Genetics, Chile). The multiple PCR products were separated on the ABI Prism 310 Genetics Analyzer (Applied Biosystems, USA) and allele calls made with GeneMapper® ID software v3.1 (Applied Biosystems, USA) by comparison with the allelic ladder provided with the amplification kit. In each round of independent analysis, a sample containing the K562 human DNA of known genotype was always used as the positive control (Promega, USA). TAAG Genetics is a laboratory that has been accredited since 2005 by the Legal and Medical Service of Chile, which is the official organism from the government that regulates the forensic and parentage testing activities with legal value in Chile. Accordingly, all procedures employed in TAAG Genetics to obtain the STR data reported here were carried out strictly following the recommendations disclosed in the 2003 manual of good laboratory practice written by the parentage testing standards program unit of the American Association of Blood Banks (AABB).

For each locus, the gene counting method was used to estimate the allele frequencies. Unbiased estimates of expected heterozygosity were computed as reported by Nei [1]. Standard error for the expected heterozygosity estimates were computed as reported by Edwards et al. [2]. Departures from Hardy–Weinberg Equilibrium (HWE) for each locus were tested using the PyPop [3] implementation of the Guo and Thompson exact test with default values. Linkage disequilibrium (LD) for pairs of loci was tested using the permutation test available with the PyPop software. Descriptive statistics such as the matching probability (pM), power of discrimination (POD), power of exclusion (PE) and polymorphism information content (PIC), were calculated as previously described [4,5]. An exact test [6] was applied to assess for genetic differentiation between the data reported in this work and (I) data reported by Jorquera and Budowle [7] and Cifuentes et al. [8] for the Chilean population, (II) data reported for Hispanic and Caucasian individuals by Butler et al. [9], (III) data reported for south American populations from Argentina (Buenos Aires) [10], Peru (Lima) [11], Colombia (Bogota) [12], Brazil (Brasilia) [13], Venezuela (Caracas) [14] and Ecuador (Quito) [15]. The software utility Struc was used for this purpose as provided in Genepop v3.4 [16] with default parameters.

The allele frequencies for the 13 CODIS STR loci over a sample of 732 Chilean individuals were obtained (Supplementary Table 1). Measures of forensic interest including the probability of match (pM), power of discrimination (POD), power of exclusion (PE) and polymorphic information content (PIC) were also calculated for each individual STR locus (Supplementary Table 2). The combined POD and PE across all loci was 0.999999999999999 and 0.9999984656, respectively, which is in full agreement with recently reported data of these descriptive statistics for the full CODIS system from the Chilean population [17]. Observed and expected heterozygosities, as well as p-values for the Guo and Thompson exact test for Hardy–Weinberg equilibrium (HWE) at each locus were computed, suggesting that all loci follow Hardy–Weinberg proportions (Supplementary Table 3). Though D16S539 locus departure from HWE was significant (p < 0.05), it was not highly significant (p = 0.039). None of the loci showed a significant departure from HWE when Bonferroni correction was applied.

To test the association of alleles at different loci, a permutation test for linkage disequilibrium between all pairs of loci was computed (Supplementary Table 4). Across all 78 pairs, only 7 pairs (D16S539-D13S317, D13S317-D8S1179, D16S539-D8S1179, D3S1358-vWA, D3S1358-TH01, D5S819-FGA and D18S51-D21S11) were significant (p < 0.05) and only one pair (D16S539-D13S317) was highly significant (p < 0.01). The results were not significant when Bonferroni correction was applied.

Two previous studies have reported allele frequencies for nine STR markers in CODIS based on Chilean samples. The first study, by Jorquera and Budowle [7], provided allele frequencies for CSF1PO,
TPOX and TH01 based on a sample of 132 individuals. The study, by Cifuentes and colleagues, reported allele frequencies for D3S1358, FGA, DBS1179, D21S11, D18S51, and D5S818 based on sample of 80 or 81 individuals, depending on the system [8]. For systems D13S317, D16S539, D7S820 and vWA, no previous reports have been published. To determine the consistency between the allele frequencies obtained in this work (TW) and those nine STR loci previously reported by Cifuentes et al. [8] (CI) and Jorquera and Budowle [7] (JO), we applied an exact test for genetic differentiation [6] over the absolute frequencies obtained by multiplying the relative frequencies with the reported sample size at each locus (Supplementary Tables 5 and 6). The test rejects the null hypothesis of no differentiation between this work and CI for FGA and between this work and JO for TH01 ($p < 0.05$). However, the difference for both loci was not found to be highly significant ($p < 0.01$). No significant differences were observed for any of the comparisons across all loci after Bonferroni correction.

Previous work by Jorquera and Budowle [7] compared Chilean population data with Hispanic and Caucasian-Americans population data relying on 10 STR markers that included 3 STR markers in the CODIS system. The results from that study suggested that Chilean allele frequencies were more closely related to Hispanic data. We applied the exact test for genetic differentiation between our data and that reported by Butler et al. [9] for Hispanic and Caucasian populations (Supplementary Table 7). When comparing our data to that for Caucasian individuals across the 13 STR loci in CODIS, all markers were highly significant ($p < 0.01$) even after Bonferroni correction [18]. In contrast, when comparing our allele frequencies to those for Hispanic individuals, we obtained seven of them to be significant ($p < 0.05$; CSF1PO, D18S51, D21S11, D5S818, D8S1179, TH01 and vWA), two of them to be highly significant ($p < 0.01$; CSF1PO and vWA) and one of them to still be statistically significant after Bonferroni correction ($p < 0.05$; vWA). Notably, for three out of the four previously unreported systems from the Chilean population (D13S317, D7S820 and D16S539) the difference between the Chilean data reported here and the Hispanic allele frequencies was not statistically significant.

Additionally, CODIS allele frequency data from six South American countries (Argentina, Brazil, Colombia, Ecuador, Peru and Venezuela) was also compared with the Chilean population data reported here (Supplementary Table 7). The results showed that Chilean allele frequencies were not closely related to most of the populations tested, with the exception of Colombian data. In this last case, we obtained that only five markers had a significant difference ($p < 0.05$; CSF1PO, D13S317, D18S51, D21S11, and D5S818) and only one marker (D5S818) to be highly significant ($p < 0.01$) and still statistically significant after Bonferroni correction ($p < 0.05$).

To this previous work, no allele frequencies were reported for loci D13S317, D7S820, D16S539 and vWA from the Chilean population. In current practice in Chile, law regulations demand that STR allele frequencies reported for Caucasian-Americans [19] must be used for these loci, for any purposes involving legal forensic procedures such as paternity testing and human identification, among others. The results obtained here show that the Hispanic or Colombian allele frequencies for STR markers in CODIS constitute a better estimate of Chilean allele frequencies than those of Caucasian-Americans or those from other populations in the following South American countries: Argentina, Brazil, Ecuador, Peru and Venezuela.

This paper strictly follows the guidelines and ISFG recommendations concerning STR nomenclature [20,21,22] and the authors declare that they accept the requirements for quality of publication of population data proposed by the journal [23].

### Appendix A. Supplementary data


### References

1. M. Nei, Estimation of average heterozygosity and genetic distance from a small number of individuals, Genetics 89 (1978) 583–590.

Ismael A. Vergara
Departamento de Genética Molecular y Microbiología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, Alameda 340, Santiago, Chile

Pamela Villouta
Sandra Herrera
TAAG Genetics S.A. Av. Andrés Bello 1245 of. 805, Tower B, Santiago, Chile
Francisco Melo\textsuperscript{a,b*}
\textsuperscript{a}Departamento de Genética Molecular y Microbiología,
Facultad de Ciencias Biológicas,
Pontificia Universidad Católica de Chile, Alameda 340,
Santiago, Chile
\textsuperscript{b}TAAG Genetics S.A., Av. Andrés Bello 1245 of. 805, Tower B,
Santiago, Chile

\textsuperscript{*}Corresponding author at:
Pontificia Universidad Católica de Chile,
Molecular Genetics and Microbiology, Alameda 340.
Facultad de Ciencias, Santiago, RM, Chile
Tel.: +56 2 686 2663; fax: +56 2 222 5515
E-mail address: fmelo@bio.puc.cl (F. Melo).

7 September 2010