

# Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

*Accepted Manuscripts* are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

# Amino Acid Composition of Human Scalp Hair as a Biometric Classifier and Investigative Lead

Ayat H. B. Rashaid<sup>1</sup>, Peter de B. Harrington<sup>1</sup>, and Glen P. Jackson<sup>2,3\*</sup>

<sup>1</sup>Center for Intelligent Chemical Instrumentation, Clippinger Laboratories, Department of Chemistry and Biochemistry, Ohio University, Athens, Ohio 45701, USA

<sup>2</sup>Department of Forensic & Investigative Science, West Virginia University, Morgantown, WV, 26506, USA

<sup>3</sup>C. Eugene Bennett Department of Chemistry, West Virginia University, Morgantown, WV, 26506, USA

**Keywords:** hair analysis, chemometrics, classification, biomarkers, amino acid analysis, FuRES.

\* To whom correspondence should be addressed.

Phone: +01 304 293 9236. Fax: +01 304 293 2663.

Email address: [glen.jackson@mail.wvu.edu](mailto:glen.jackson@mail.wvu.edu)

## Abstract

Hair shaft analysis is becoming increasingly important in several applications of forensic sciences. Keratin is the key component of human scalp hair and is composed of all 21 known amino acids, albeit in very different proportions. The method for the amino acid determination included protein acid hydrolysis followed by trimethylsilyl (TMS) derivatization of the amino acids, and the subsequent quantitation using by gas chromatography/mass spectrometry (GC/MS). Amino acid composition of scalp hair of 64 Jordanian subjects (33 males and 31 females) with ages ranging from 1 and 77 years have been analyzed. Statistical comparisons between classification groups were based on the abundance of 14 abundant and acid-resistant amino acids, and included classification of hairs with a fuzzy rule building system (FuRES). Using leave-one-individual-out cross validation, the FuRES classification rate was 94% for sex, 83% for age group, and 61% for the region of origin. For predicting sex from amino acid concentration in hair, the essential amino acids Phe and Thr gave the most significant differences with respect to their  $F$  statistic (i.e., ratio of between-group to within-group variation), so they are the most discriminating for sex. Based on the same hair analyses, the non-essential amino acids Gly and Ala provide the largest loading scores classifying the subjects into two arbitrary age groups, <44 and >44 yrs. For region of origin classification, the amino acids Cys and Tyr had the highest loading scores in the classification rules and were therefore most discriminating. The techniques developed through this paper could complement the current methods of hair analysis, which include physical examinations and genomic or mitochondrial DNA analysis.

## Introduction

Human hair has been a subject of intensive investigation for a number of years. Study of the hair shaft is becoming increasingly important in several applications of forensic sciences.<sup>1,2</sup> In addition, hair is one of the more robust non-skeletal tissues in decomposed bodies and postmortem cases of forensic interest.<sup>2</sup> The growth rate of human scalp hair ranges from 0.8 to 1.3 cm per month.<sup>3-5</sup> Hair consists of the cuticle, cortex, and medulla which are formed via the keratinization of dying cells growing out of the root. Hair is mainly composed of alpha keratin protein (65-95%), and the remaining constituents are water, lipids, pigment, and trace elements.<sup>4,6-8</sup> Keratin contains strong disulfide bonds that link adjacent keratin chains, and contains intramolecular bonding of polar and non-polar amino acids, which lead to a stable structure that is highly resistant to chemical and biological degradation. The amino acid composition of human hair keratin may vary significantly among individuals without exhibiting obvious phenotypic differences.

Keratin, the protein which is the key component of hair, is composed of 21 known amino acids.<sup>9</sup> Changes in the amino acid composition of hair have long been a target of interest to nutritional scientists and health scientists. Profiling amino acids in human hair could be utilized for characterization of various types of hair derived from different individuals.<sup>3,10</sup> Therefore, the amino acid profile of human hair could be a useful tool with regard to classification of individuals into groups of biological or behavioral traits, in a complementary manner to work using isotope ratio analysis of amino acids in human hair.<sup>11</sup>

Hair follicles undergo a growth cycle comprising three stages: the anagen stage is the active growing phase where the cells in the bulb of the follicle are rapidly dividing; scalp hair stays in this phase from 4 to 6 years. The catagen stage, which is the transition phase, lasts for 2-

1  
2  
3 3 weeks. During this phase cell division stops and the hair shaft becomes fully keratinized. The  
4  
5 telogen stage is the resting phase wherein hair growth stops completely. The duration of this  
6  
7 stage is from 4 to 6 months.<sup>4, 9, 12</sup> Nuclear DNA profiling results are typically only successful on  
8  
9 hair roots in an appropriate growth cycle (anagen or catagen), and when a subject's nuclear DNA  
10  
11 is already in the reference database.<sup>13</sup> However, most hair found at crime scenes has roots that  
12  
13 are in the telogen or resting phase,<sup>14, 15</sup> wherein nuclear DNA is in very low abundance and  
14  
15 severely degraded, and databases only contain DNA profiles from a small percentage of the  
16  
17 population.  
18  
19  
20

21  
22 Mitochondrial DNA profiling can be performed on the hair shaft, but the accompanying  
23  
24 statistics are far less powerful than nuclear DNA because it is inherited maternally.<sup>16</sup>  
25  
26 Accordingly, when DNA is not recoverable from hair samples, or when reference DNA samples  
27  
28 are not available, analysis of the amino acid content of shaft hair could be a reliable method for  
29  
30 linking questioned hair at the crime scene with suspects or victims, or for determining a  
31  
32 biometric traits and investigative leads about a donor.  
33  
34  
35

36  
37 The amino acid composition of hair is known to be influenced by a number of factors  
38  
39 including age,<sup>8, 17-20</sup> sex,<sup>18, 21</sup> hair color,<sup>18, 22</sup> and race,<sup>22-24</sup> as well as dietary habits and  
40  
41 geographic origin.<sup>9, 17, 25, 26</sup> In the context of investigating the relationship between age and  
42  
43 amino acids of human hair, a decrease in tryptophan (Trp) concentration was noted between age  
44  
45 groups 2-5 years and 13-19 years.<sup>8</sup> Furthermore, a gradual increase in Trp was observed  
46  
47 between the ages of 13-19 and 60-80 years.<sup>19</sup> Statistically significant changes with increasing  
48  
49 age showed increasing concentrations of serine (Ser), arginine (Arg), and proline (Pro), whereas  
50  
51 a decrease in concentration of aspartic acid (Asp), glycine (Gly), alanine (Ala), valine (Val),  
52  
53 isoleucine (Ile) and lysine (Lys) were shown with increasing age.<sup>8</sup> It has also been observed that  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 cystine (Cyt) and cysteine (Cys) content decreased with age. For example, higher values of Cyt  
4 amino acid was found among infants (9 months) compared with children (8-9 years).<sup>17</sup>  
5  
6  
7  
8 However, both Clay *et al.*<sup>18</sup> and Wilson and Lewis<sup>21</sup> concluded that no relation can be  
9 demonstrated between Cys and Cyt contents of the hair and age of the subjects.  
10

11  
12  
13       Investigators have also examined the effects of ethnicity and sex on the amino acid  
14 composition in human hair. Regarding sex differences, Clay *et al.* found significantly more Cyt  
15 and Cys in the hair of the males compared to the females.<sup>18</sup> Rutherford and Hawk<sup>22</sup> also found  
16 that the Cyt and Cys content of Caucasian males was greater than that of females. However, in  
17 related work, Wilson and Lewis<sup>21</sup> analyzed twenty-three samples of hair from children and six  
18 samples from adults and found no consistent relationship between sex and the content of Cys and  
19 Cyt. Panayiotou<sup>24</sup> reported that females have a higher content of glutamic acid (Glu), Asp, and  
20 Trp amino acids than males.  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

31  
32       A limited number of studies have investigated the amino acid composition of scalp hair  
33 under different dietary habits of different regions of origin. Takanohashi<sup>17</sup> conducted an  
34 experiment to determine whether the Cys content in the hair of infants and children could be  
35 changed under various circumstances i.e., deficiency of vitamin A, animal protein, and other  
36 nutrients in their diet. Hair samples were obtained from three different districts; region A  
37 (people living on Japanese barnyard millet and very little fish), region B (people living on rice  
38 and moderate amount of fish), and region C (people living on a well-balanced diet). Results  
39 indicated that the Cys content of hair was low when the subject's diets were deficient in vitamin  
40 A or in high quality of protein, while it was observed that by supplementing their diets with liver  
41 oil or with plenty of milk, the Cys content increased. These results are in agreement with  
42 observations of experiments conducted on rats.<sup>26</sup> Heard and Lewis<sup>26</sup> reported that the addition of  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 supplementary methionine (Met) to the basal diet known to be deficient in its content of sulfur  
4  
5 containing amino acids resulted in the production of hair with a higher percentage of Cys than  
6  
7 the hair of animals receiving the basal un-supplemented diet. Moreover, the experiments of  
8  
9 Lightbody and Lewis<sup>25</sup> clearly show that the Cys content of hair can be changed by diet. They  
10  
11 recognized a lower Cys content in the hair of rats than in those of rats fed on diets high in Cys  
12  
13 content.  
14  
15

16  
17 The essential amino acids of keratin that cannot be synthesized from other substances in  
18  
19 the human body (i.e., de novo synthesis) are leucine (Leu), Ile, Thr, Trp, Met, Lys, Phe, and Val.  
20  
21 These amino acids are normally furnished in the diet in the form of protein (e.g., meat, milk and  
22  
23 eggs).<sup>9</sup> Therefore, dietary habits could be an important variable for the characterization of the  
24  
25 essential amino acids in human hair.  
26  
27

28  
29 The previous literature provides some evidence to suggest that sex, age, and region of  
30  
31 origin may produce variation in certain amino acid contents of human hair. However, relatively  
32  
33 few extensive studies of this nature have been reported. Profiling amino acids in human scalp  
34  
35 hair has never been studied in the Arab World, and, to our knowledge, multivariate classification  
36  
37 of individuals by their hair amino acid profiles has not been demonstrated before. Jordan has  
38  
39 been selected for this purpose because it has three distinct geographical regions (mountain,  
40  
41 valley, and desert) with distinct dietary habits for each region. This study is the first to show the  
42  
43 feasibility of the amino acid profiles to classify subjects based on sex, age, and region of origin.  
44  
45

46  
47 With reference to the previous studies, the general statement seems valid that relations  
48  
49 between the amino acid distribution and each of sex, age, and region of origin were not  
50  
51 established because of the number of samples was typically less than 20.<sup>8, 17, 22, 27</sup> In the present  
52  
53 study, the scalp hair from 64 subjects spanning different age groups from three different regions  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 of Jordan (mountain, Jordan valley, and Bedouin region) with adequate information about the  
4  
5 subjects (age, sex, height, weight, hair color, medical history, dietary habits, and chronic or acute  
6  
7 diseases) were studied to establish the feasibility of using amino acid profiles of hair for  
8  
9 classification of individuals. Moreover, the number of subsamples (3 trials) for each subject  
10  
11 provided greater confidence for the method and the statistical tests.  
12  
13  
14

15 Previous work has shown that Trp, Cyt, Met, Asn, Gln, His, and Arg amino acids are  
16  
17 detrimentally affected by acid hydrolysis.<sup>28</sup> For this reason, this study focuses on the 14 most  
18  
19 abundant amino acids in the protein of human hair, as obtained by acid hydrolysis using 6 M of  
20  
21 hydrochloric acid (HCl).<sup>28-31</sup> The amino acids were quantified via their trimethylsilyl (TMS)  
22  
23 derivatives<sup>32-35</sup> using gas chromatography/mass spectrometry (GC/MS).<sup>29, 36, 37</sup> The main  
24  
25 objective of this study is to provide a discriminating technique for distinguishing between hair  
26  
27 samples from different human subjects based on factors related to sex, age, dietary habits, and  
28  
29 region of origin that have been assumed to contribute to variations in the amino acid contents of  
30  
31 human scalp hair.  
32  
33  
34  
35

## 36 **Methods and instrumentation**

### 37 **Hair samples**

38  
39 The collection of hair samples for this study was approved by the Ohio University  
40  
41 institution review board (IRB# 12X029) and King Abdullah University hospital institution  
42  
43 review board (IRB# 10/215/2444). Scalp hair from 64 healthy Jordanian individuals was  
44  
45 collected randomly from three different geographical regions of Jordan, each with a distinct  
46  
47 geography, culture, and dietary habits. People who live in each region have their own nutritional  
48  
49 habits. For example, the Bedouin, who live in deserts, typically depend on dairy products and  
50  
51 lamb meat. Residents of the mountainous region depend more on fruits, vegetables, and rice.  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 The primary food of people in the valleys is fish. Jordan therefore offers a unique opportunity to  
4 study people with quite distinct geographic and dietary lifestyles.  
5  
6

7  
8 The names, addresses and phone numbers of potential subjects were taken from a  
9 database of the General Civil Status department. This department is a national bureau that stores  
10 all the personal information about all Jordanians. Pregnant women and subjects who were  
11 diagnosed with any chronic or acute disease were excluded from the study. The subjects were 33  
12 men and 31 women between the ages of 1 and 77 years. They were also classified according to  
13 three geographical regions: 29 from the mountain, 23 from the valley, and 12 from the eastern  
14 desert regions.  
15  
16  
17  
18  
19  
20  
21  
22  
23

24  
25 In the scheduled home visits, consent forms from adults and assent forms from minors  
26 were signed prior to filling out interview questionnaires regarding age, sex, biometric  
27 information, health status, hair care, cosmetic treatment, and dietary habits. A pencil-diameter (1  
28 cm) lock of hair adjacent to the scalp was cut using surgical scissors from the posterior vertex  
29 (back of the head) of each subject. All samples were coded and no specific individual was  
30 identifiable thereafter. All hair used in this study was reported to be rarely altered chemically,  
31 i.e., bleaching, chemical straightening, dyes, and permanent waves. Because of Jordan's cultural  
32 factors, female hair was never directly exposed to the sun and none of the subjects smoked  
33 regularly or drank alcohol. These factors may not transcend different populations or study  
34 groups.  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

#### 48 **Chemicals and supplies**

49  
50 Any variation in conditions from one run to the next was controlled by referencing all  
51 data to an internal standard that was added to each acid hydrolysate.<sup>38</sup> L-norvaline (Nor) (Sigma-  
52 Aldrich, St. Louis, MO, USA) was used as an internal standard because it is an exogenous amino  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 acid that is not synthesized by the human body. The derivatizing agent was *N, O*-  
4 bis(trimethylsilyl)trifluoroacetamide (BSTFA) (Supelco Analytical, Bellefonte, PA, USA). A  
5  
6 solution of 6 M HCl was used as a hydrolysis agent to liberate free amino acids from the hair  
7  
8 shaft. Acetonitrile, methanol, acetone, and chloroform were purchased from GPS Chemicals  
9  
10 (Columbus, OH, USA).  
11  
12

13  
14  
15 Glass vials of 4-mL volume with a phenolic rubber lined cap (Qorpak, Bridgeville, PA,  
16  
17 USA) were used in this study. Solutions were filtered through a 13 mm × 0.45 μm,  
18  
19 polyvinylidene difluoride (PVDF) filter (Bonna-Agela Technologies, Wilmington, DE, USA).  
20  
21 The nitrogen generator was purchased from Parker Hannifin Corporation (Haverhill, MA, USA).  
22  
23

#### 24 **Hair analysis**

25  
26  
27 Approximately 1 g from each hair specimen was rinsed twice in deionized water and  
28  
29 vortexed in 1:1:1 (by volume) solution of chloroform, methanol, and acetone for 1 min to  
30  
31 remove surface contamination, external oils, and lipids.<sup>39</sup> The hair was then dried under a  
32  
33 constant stream of 4 L/min of nitrogen to remove the excess reagents. Afterwards, the hair was  
34  
35 cut into 1 cm lengths and approximately 20 mg of hair was physically degraded and  
36  
37 homogenized by placing the hair into a capped plastic vial with 8 stainless steel beads and  
38  
39 pulverized for 1 min using a mini bead beater (BioSpec Products Inc., Bartlesville, OK, USA) at  
40  
41 4800 rpm. Because there is no known hair standard available with which to validate the method,  
42  
43 horse heart myoglobin was used. Although myoglobin does not contain disulfide bonds, it does  
44  
45 contain a high degree of alpha-helical structure; much like the keratins found in hair. A report on  
46  
47 the method development and optimization is provided elsewhere.<sup>28</sup>  
48  
49  
50  
51

52  
53 Determining the amino acid content of human hair involved the acid hydrolysis in 6 M  
54  
55 HCl until its individual amino acids separate and become available for detection. The hair  
56  
57  
58  
59  
60

1  
2  
3 powder from each sample was separated into three subsamples and weighed to  $3.00 \text{ mg} \pm 0.03$   
4 mg (1 s.d.) and transferred into a 4-mL glass vial with a phenolic rubber lined cap. Thereafter,  
5  
6 mg of Nor was added. Each subsample was hydrolyzed by adding 0.3 mL of 6 M HCl to the  
7  
8 vial and was tightly capped. To heat the samples, an aluminum block with holes to  
9  
10 accommodate the glass vials was mounted on a hot-plate for 24 h at  $110^\circ\text{C}$ .<sup>29-31, 40</sup> After cooling  
11  
12 the solution to room temperature for 5 min, the solutions were filtered and dried under a constant  
13  
14 4 L/min nitrogen stream because the existence of excess moisture can result in poor reaction  
15  
16 yield and instability of the derivatized products. The three subsamples of hair from the same  
17  
18 individual were digested across spans of 3 days. In some cases, the amount of collected hair  
19  
20 from an individual was not sufficient to prepare three digestions, so they were done in duplicate.  
21  
22  
23  
24  
25  
26

27 The dried amino acid residues were dissolved in a 100- $\mu\text{L}$  aliquot of acetonitrile and  
28  
29 derivatized with the addition of a 100- $\mu\text{L}$  aliquot of BSTFA.<sup>32, 38, 41</sup> The glass vial was tightly  
30  
31 capped, ultrasonicated for 1 min and mounted on a hot-plate for 30 min at  $100^\circ\text{C}$ .<sup>42, 43</sup> After  
32  
33 cooling the solution at room temperature for 10 min, a 100- $\mu\text{L}$  aliquot of the solution was  
34  
35 transferred into a 3-mL glass auto-sampler vial containing a 200  $\mu\text{L}$  plastic insert.  
36  
37  
38

### 39 **Instrumentation**

40  
41 The derivatized amino acids were analyzed using GC/MS and identified according to  
42  
43 their retention time<sup>44, 45</sup> and EI fragmentation patterns using independently analyzed BSTFA  
44  
45 derivatives of the amino acid standards. Amino acid derivatives were analyzed with a Shimadzu  
46  
47 GC/MS instrument (QP-2010SE, Scientific Instrument, Inc. Columbia, MA, and USA). Amino  
48  
49 acids were separated on 5% diphenyl-dipolymethylsiloxane capillary column (SHRX1-5MS, 30  
50  
51 m  $\times$  0.25 mm  $\times$  0.25  $\mu\text{m}$ ). Ultra high purity Helium was used as a carrier gas at a constant flow  
52  
53 rate of 21.5 mL/min. The transfer column flow was 1.0 mL/min. The temperature of the column  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 was programmed to rise from 70 °C to 170 °C at a rate of 10 °C/min, and then was ramped to 280  
4 °C at a rate of 30 °C/min, at which it was held for 3 min. The total run time was 16.6 min.

5  
6  
7  
8 Sample injection was performed in split injection mode (1:20 ratio) at 280 °C using an injection  
9 volume of 1 µL. Triplicate injections of triplicate samples were run in random block design  
10 using acetonitrile as a solvent blank before each injection. Three solvent vials of acetonitrile,  
11 methanol, and acetone were used as cleaning solvents for the autosampler injection syringe.  
12  
13  
14  
15

16  
17 The mass spectrometer was operated in Fast Automated Scan and SIM Type mode  
18 (FASST), which switches between full scan mode and selected ion monitoring during each GC  
19 analysis. The mass spectrometer was operated in full scan mode (TIC) from  $m/z$  of 50 to  $m/z$  of  
20 500 with a scan time of 0.3 s, and selected ion monitoring mode (SIM) with a scan time of 0.2 s.  
21  
22 For SIM, the appropriate ion set of one target ion and two reference ions as a characteristics mass  
23 fragment of the derivatized amino acid was used (Table 1). Selection of the characteristic ions  
24 for the quantification ion is very important. The target ions are the fragment ions of the highest  
25 intensity (base peak), while the reference ions are the two fragment ions of the next greatest  
26 intensities.<sup>28, 42</sup> Whereas the TIC shows partial baseline overlap for Ile and Pro, the selected  
27 quant ions enabled sufficient discrimination between the two amino acids to enable unambiguous  
28 peak integration and quantitation.  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42

#### 43 **Data analysis**

44  
45  
46 Statistical analyses were performed using MATLAB R2012b (The MathWorks Inc.,  
47 Natick, MA). All statistical tests were conducted at the 95% confidence level.  
48  
49

50  
51 Amino acids in hair protein were determined by the measurement of their relative peak  
52 areas measured by GC/MS that were referenced to the peak area of norvaline internal standard.  
53  
54  
55 In previous work,<sup>28</sup> selected ion monitoring (SIM) yielded the best performance. The  
56  
57  
58  
59  
60

1  
2  
3 concentration ( $\mu\text{mol/mL}$ ) for each amino acid was calculated using external calibration curves of  
4  
5 liquid standards. A random block design was used to collect triplicate measurements of each  
6  
7 hair digestion across spans of 2 days.  
8  
9

10 A variety of statistical techniques were used to correlate or discriminate individuals or  
11  
12 groups of individuals based on amino acid compositions. For example, to analyze the  
13  
14 multivariate data, a multivariate fuzzy rule-building expert system (FuRES)<sup>46</sup> was used to build  
15  
16 classifier and reveal the characteristic amino acids for the classes. FuRES constructs a  
17  
18 classification tree that allows the visualization of the inductive structure of the classes. The  
19  
20 inductive classification tree is a collection of membership functions for which each branch is a  
21  
22 multivariate fuzzy rule. The entropy of the system decreases as the classification tree branches  
23  
24 until all the objects are grouped into a single class at the leaves of the tree.<sup>46</sup> FuRES has been  
25  
26 compared with more traditional classification methods in a variety of different analytical  
27  
28 applications.<sup>47-55</sup>  
29  
30  
31  
32  
33

34 Because FuRES is a supervised classifier, the results were validated using leave-one-out  
35  
36 cross-validation. Several approaches were used for validation. Leave-one-out cross-validation  
37  
38 involves leaving out a measurement (or profiles), constructing the FuRES rules based on the  
39  
40 remaining measurements (or profiles), and then classifying the omitted measurement (or profile).  
41  
42 The classification success is repeated for every measurement (or profile) to obtain an overall  
43  
44 estimate of the ruggedness of the classification system. The FuRES discriminant weights are  
45  
46 stored for each model that is built during this process and the average weight with an arbitrary  
47  
48 95% confidence interval can be used as a diagnostic to determine which amino acids are most  
49  
50 characteristic for a specific the classification. Note that because a random block design was used  
51  
52 in acquiring the data, that replicate amino acid profiles were typically collected on different days.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Besides leaving each profile out, another mode of validation averaged 3 the amino acid  
4 profiles for each separate hair digestion. The averages were classified and each averaged amino  
5 acid profile corresponded to a separate digestion. Lastly, the averages were calculated for each  
6 triplicate and the three digestions, so each average amino acid corresponded to a unique  
7 individual. This test is the most stringent because the model was built by leaving out profiles  
8 from an individual as opposed to one of several replicates measurements of an individual. Using  
9 averaged-within-person profiles is therefore the most honest representation because each profile  
10 is truly an independent measure of the classification factor.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21

22 The accuracy of the tests were evaluated by their sensitivity and specificity of the  
23 classifier predictions.<sup>56</sup> Sensitivity is defined as the ratio of positive classifier responses for  
24 positive subjects. Specificity is defined as the ratio of negative classifier responses for negative  
25 subjects. Sensitivity plotted with respect to the false positive rate (1-specificity) can be  
26 displayed graphically using the receiver operating characteristic (ROC) curve.<sup>56</sup> The receiver  
27 operating characteristic curve (ROC) was generated to evaluate the sex, the age group, and the  
28 region of origin classifiers. The accuracy of a classifier depends on its ability to maintain the  
29 sensitivity and the specificity, so that the ideal result will have an area under the curve (AUC) of  
30 1.<sup>56</sup> The area under the curve is calculated using the trapezoidal integration function of  
31 MATLAB.  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45

46 Analysis of variance (ANOVA) is a statistical technique for determining the significance  
47 of the differences between two or more groups; it was used to compare the concentration  
48 ( $\mu\text{mol/mL}$ ) of each amino acid in different groups of subjects and to find which amino acids  
49 have the largest ratio between within-person variance and between-person variance.<sup>57, 58</sup> Based  
50 on the ANOVA results of region of origin groups, post-hoc statistical tests such as Tukey's  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 honest significant difference (HSD) was used to identify which groups were significantly  
4  
5 different from another. Principal component analysis (PCA) was used to reduce the number of  
6  
7 variables in a data set while retaining as much variation from the original data set. PCA was also  
8  
9 used to visualize the clustering of the subjects scores.<sup>59</sup> ANOVA-PCA is a method that  
10  
11 combined ANOVA with PCA.<sup>60</sup> Pooled ANOVA-PCA<sup>61, 62</sup> was used to evaluate the differences  
12  
13 among the class means for the three regions of origin, sex, and age after averaging the 9 profiles  
14  
15 for the 64 individuals.  
16  
17  
18  
19

## 20 **Results and discussion**

21  
22 The 14 amino acids of interest in this study are Ala, Gly, Val, Leu, Ile, Pro, Ser, threonine  
23  
24 (Thr), Asp, Cys, Glu, Phe, Lys, and Tyr. Sharp single peaks for all amino acids were obtained  
25  
26 except for Gly, which gave both di-trimethylsilyl derivative and tri-trimethylsilyl derivatives that  
27  
28 were obtained after BSTFA derivatization reaction.<sup>28, 35</sup> See Figure 1. The observation of two  
29  
30 products for certain amino acids, although not ideal, does not necessarily prevent quantitation. If  
31  
32 both products are chromatographically resolved, the sum of the two peak heights or peak areas  
33  
34 can be used for quantitation.  
35  
36  
37  
38

39 FuRES was used to evaluate the ability to classify the subjects according to the sex, age  
40  
41 group, and region of origin based on multivariate analyses of the abundance of 14 amino acids in  
42  
43 human hair. The age group was arbitrarily defined as two groups: one containing 1-49 year-olds  
44  
45 and the other group containing 50-77 year-olds. FuRES was applied to the 528 amino acid  
46  
47 profiles that for the most part were obtained from (3 replicate measurements of 3 separate  
48  
49 digestions) for each individual. However, for some individuals the quantity of hair that was  
50  
51 collected was not sufficient for 3 digestions, so 48 profiles were not collected. In a similar  
52  
53 manner, FuRES was also applied to the 177 average amino acid profiles of the three injections  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 for each digestion. In the third experiment, FuRES was applied to the average of 9  
4  
5 measurements of each individual (average across individuals). The classifier prediction rates of  
6  
7 sex, age, and region of origin after leave-one-out cross-validation are presented in Table 2.  
8  
9

10 Using leave-one-profile-out cross-validation, all three factors could be correctly predicted  
11  
12 at better than 81% success rate when using all the objects (measurements) to build the  
13  
14 classification models. When averaging across individuals, sex had the best classification rate of  
15  
16 94% compared to 83% for age and 61% for region of origin. The relatively poor performance  
17  
18 for the classification by region of origin likely arises from the weakness of this factor relative to  
19  
20 other compounding factors like age and sex, and the fact that origin was separated into three  
21  
22 groups instead of two groups for the age and sex factors.  
23  
24  
25

### 26 27 **Sex classification**

28  
29 A FuRES classification tree is the collection of rules used to separate all the profiles. The  
30  
31 rules are obtained by minimizing the entropy of classification ( $H$ ). The root/trunk of the tree  
32  
33 (typically shown with the root at the top) contains the most general rule. Moving down the tree  
34  
35 from the root to the leaves at each branch the rules gets more specific (i.e., they separate objects  
36  
37 that are closer together in the data space). Figure 2 is a FuRES classification for sex using the  
38  
39 average profiles for individuals. This three-rule tree yielded a 94% classification rate, whereas  
40  
41 Panayiotou<sup>24</sup> reported 80% success rate of fuzzy clustering for sex using FT-IR.  
42  
43  
44

45  
46 Figure 3 is the ROC curve that was generated to identify the sex of the individuals. The  
47  
48 true-positive rate or sensitivity is plotted on the ordinate as a function of the false positive rate on  
49  
50 the abscissa. The classification threshold decreases as the curve progress from left to right.  
51  
52 Accuracy of the classifier is measured by the area under the curve (AUC). The ideal diagnostic  
53  
54 test is one that reaches the upper left corner of the graph with the AUC equal to 1.<sup>56</sup> The area  
55  
56  
57  
58  
59  
60

1  
2  
3 under the curve of Figure 3 is 0.95 indicating very good performance, especially for a classifier  
4 that is generalizing for individuals, which indicates that hair samples from new individuals are  
5 likely to be successfully classified.  
6  
7  
8  
9

10 The loadings of the FuRES discriminant for rule #1 for sex classification is given in  
11 Figure 4, which separates the subjects into thirty-three males and three females on one side and  
12 twenty-eight females on the other side. Two more rules (loading scores not shown) are required  
13 to distinguish the three remaining females from the twenty-eight males. The positive peaks in  
14 Figure 4 correspond to amino acids that are loaded higher for females than males (Val, Pro, Ser,  
15 Glu, and Phe), and therefore more abundant in females than males. Negative peaks correspond  
16 to amino acids that are loaded heavier for males than females (Ala, Gly, Leu, Ile, Thr, Asp, Cys,  
17 Lys, and Tyr), and therefore more abundant in males than females. If the 95% confidence  
18 intervals of the error bars cross the zero line, the amino acid is not significant for the  
19 classification. Phe is clearly the most prominent feature and is found to be present in higher  
20 abundance in females relative to males. In contrast, males typically had higher concentrations of  
21 Thr and Lys. Although less prominent, males were found to have higher Cys content than  
22 females, in agreement with the results of others.<sup>18, 22</sup>  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40

41 As expected from the FuRES results, a significant difference of ( $p < 0.01$ ) was found  
42 between within-person variances and between-person variances for the two sex groups after  
43 ANOVA-PCA analysis. The impact of sex and source of variation between the groups was  
44 evaluated by comparing the concentration ( $\mu\text{mol/mL}$ ) of each amino acid in both male and  
45 female groups (Table 3). In Table 3, the essential amino acids (Phe, Ile, Val, and Thr) and the  
46 non-essential amino acids (Pro and Glu) have significant differences for the ratio of their  
47 between-group to within-group variance to variance (i.e.  $F$  statistic). Therefore, these amino  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 acids should be highly discriminating for sex. The loading scores in FuRES also indicate that  
4 these amino acids are heavily loaded and therefore very important for classifying sex.  
5  
6

7  
8 Consequently, the multivariate FuRES and the univariate ANOVA results agree.  
9

10 The results of the ANOVA and FuRES agree with the results obtained by Panayiotou,<sup>24</sup>  
11 who reported that Phe and Glu amino acids correlate with the proportion of  $\alpha$ -helical structure  
12 relative to  $\beta$ -sheet structure in hair. Panayiotou found that females have more  $\alpha$ -helix structure in  
13 their hair than males and therefore have more Glu and Phe. In this present study, Phe has the  
14 most significant difference for the within-group variance to between-group variance and the  
15 heaviest loaded peak in the female group. In our population, this result could also be explained  
16 by ultraviolet radiation (UV) that has the ability to decrease the Phe amino acid of hair via photo-  
17 degradation upon exposure to sunlight.<sup>9</sup> A response in the survey indicated that all of the  
18 females subjects in the present study covered their hair with a Hijab that presumably blocks UV  
19 from the sun and prevents photodegradation of Phe. This result has also been supported by  
20 Nogueira *et al.*,<sup>24</sup> who reported that the aromatic amino acids such as Phe were degraded upon  
21 exposure to sunlight.  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37

38  
39 Conflicting results were found with relation to the impact of sex on the contents of Cys in  
40 human hair. Clay *et al.*<sup>18</sup> and Rutherford and Hawk<sup>22</sup> found significantly more Cys in the hair of  
41 males compared to females. In contrast, Wilson and Lewis<sup>21</sup> found no consistent relation  
42 between the Cys and sex. In the present study, no such difference was found in Cys  
43 concentration between males and females.  
44  
45  
46  
47  
48  
49  
50  
51  
52

### 53 **Age classification**

54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 To examine the effect of age on the amino acid content, the subjects were split into two  
5  
6 arbitrary age groups: group A comprised subjects with ages from 1 to 49 years and group B of  
7  
8 elder people with ages ranging from 50 to 77 years. Classification of subject groups was modeled  
9  
10 by FuRES. The two groups were separated using a four-rule tree. For example, the second rule  
11  
12 separates 47 subjects of group A and one subject of group B. The fourth rule separates 9  
13  
14 subjects of group B and only one subject of group A. Figure 5 is a FuRES classification tree  
15  
16 demonstrating the classification model. The correct classification rate for the model was 83%, as  
17  
18 assessed using the leave-one-individual-out cross-validation.  
19  
20

21  
22 The area under the curve in Figure 6 is 0.85, representing a very good test that could  
23  
24 effectively distinguish age group A from age group B. The ROC is also supported by the similar  
25  
26 FuRES classification rate of 83%.  
27  
28

29  
30 A significant difference of ( $p=0.03$ ) was found between within-person variances and  
31  
32 between-person variances for the two groups after ANOVA-PCA analysis. The sources of  
33  
34 variation among the groups were evaluated by comparing the average concentrations ( $\mu\text{mol/mL}$ )  
35  
36 of each amino acid from the three regions are presented in Table 4. ANOVA revealed that the  
37  
38 non-essential amino acids (Gly and Ala) and essential amino acids (Thr and Asp) appear to be  
39  
40 the most discriminating between age groups;  $F$  values were 23, 10, 8, and 6, respectively.  
41  
42

43  
44 The findings indicated that the concentration of Ala, Gly, Ile, Thr, and Asp are  
45  
46 significantly higher in the younger group A, whereas, Gly, Phe, and Cys concentrations are  
47  
48 significantly higher in the older group B. The lower abundance of Asp in group B (older group)  
49  
50 is supported by the finding by Rieck,<sup>8</sup> who concluded that the concentration of Asp significantly  
51  
52 decreases with age. On the contrary, our finding of higher content of Cys in group B (older  
53  
54 group) does not match the findings of Takanohashi,<sup>17</sup> who reported a decline in Cys  
55  
56  
57  
58  
59  
60

1  
2  
3 concentration with increasing age. Wilson and Lewis<sup>21</sup> and Clay *et al.*<sup>18</sup> found no relation  
4  
5 between Cys content of hair and age. Sex did not correlate significantly with age.  
6  
7

8 No statistically significant difference was found between age groups for Ser. Rieck<sup>8</sup>  
9  
10 studied the age-dependent hair of three sisters between the ages 1 to 24 years and reported  
11  
12 increasing Ser concentration with increasing age. However, our age groups reported here span a  
13  
14 much larger range of years, so the within-group variation for the ‘young’ group A might obscure  
15  
16 potential differences between the older group B .  
17  
18

19  
20 Whereas dietary habits could be an important variable that affect the characterization of  
21  
22 the essential amino acids in human hair, genetic or phenotypic factors are more likely to  
23  
24 influence the non-essential amino acids. Relatively significant variation in Thr and Ile between  
25  
26 the two age groups could be attributed to their dietary habits.  
27  
28  
29  
30  
31

### 32 **Region of Origin Classification**

33  
34 Figure 7 is a FuRES classification tree demonstrating the classification model of the  
35  
36 region of origin. The classification rate for the model was 61% after the leave-one-individual-  
37  
38 out cross-validation. The three regions of origin were separated using a nine-rule tree. For  
39  
40 example, the second rule separates 22 subjects of the mountain region to the right and one  
41  
42 subject from the desert to the left side of the tree.  
43  
44

45  
46 In the ROC curves, the AUCs are 0.72, 0.71, and 0.83 for the mountain, the valley, and  
47  
48 the desert, respectively (Figure 8). The classification according to region of origin is not nearly  
49  
50 as accurate as by sex or by age group, but it also is not random that is indicated by an AUC of  
51  
52 0.5. Region of origin is less effective than age and sex in discriminating the hairs of individuals.  
53  
54  
55 A larger and more-equal sample size of each region could improve the classifier.  
56  
57  
58  
59  
60

1  
2  
3 ANOVA-PCA of the 14 amino acids gave a significant difference among the means of  
4 the three groups of amino acid profiles for region of origin ( $p$  of 0.03). The sources of variation  
5 among the groups were evaluated by comparing the average concentrations ( $\mu\text{mol/mL}$ ) of each  
6 amino acid from the three regions (Table 5). The amino acids Cys, Ile, Phe, and Thr have the  
7 most significant differences for between-group variance to within-group variance. Therefore,  
8 they may be important for the classification of subjects according to their region of origin.  
9

10  
11 Cys is a semi-essential amino acid, which means that it can be biosynthesized in humans  
12 and furnished by diets.<sup>9</sup> Cys is found in most high-protein foods, including meat, eggs, dairy  
13 products.<sup>17</sup> Higher levels of Cys in the hair samples from desert region compared to mountain  
14 and valley region is an expected observation, which may be attributed to the preference of meat,  
15 egg, and dairy product by the subjects living in the deserts of Jordan. This outcome is  
16 compatible with the findings of Takanohashi,<sup>17</sup> who observed that subjects who drink plenty of  
17 milk have higher Cys content in their hair than those subjects who consume less milk. In  
18 addition, the Bedouin, who reside in the desert region, are exposed to the sun more than other  
19 groups because they live in tents and tend sheep. They reported in the study that their scalp hair  
20 was exposed to the sunlight more than 4 hrs per day. Ultraviolet radiation (UV) has the ability to  
21 alter the amino composition of hair in photo-degradation upon exposure to sunlight.<sup>9</sup> Cyt  
22 contains two Cys amino acids that are linked by a disulfide bond.<sup>3</sup> Breaking the disulfide bond  
23 through photochemical degradation of Cyt produces more Cys residues.<sup>9, 63</sup>  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47

48 The abundance of Phe and Tyr amino acids are significantly lower in the desert group  
49 and this result is similar to the finding of Nogueira *et al.*,<sup>63</sup> who reported that the aromatic amino  
50 acids such as Tyr and Phe were degraded upon exposure to sunlight. The significant variation  
51 between the three groups in essential amino acids (Ile, Leu, Phe, and Lys) plays an important  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 role in the classification of region of origin. It may be attributed to a combination of the  
4  
5 different dietary habits in the three different regions and sunlight exposure related to lifestyle  
6  
7 habits.  
8  
9

## 10 **Conclusions**

11  
12  
13 The ability to classify the hair of subjects into age, sex, and region of origin using profiles  
14  
15 of 14 amino acids obtained from hair samples has been demonstrated for a large-scale and  
16  
17 rigorous evaluation. The classification rules extracted from FuRES were consistent with the  
18  
19 results of ANOVA, and for the most part other reports in the literature. The findings of this  
20  
21 study provide good evidence for the feasibility of amino acids to distinguish between subjects or  
22  
23 groups of subjects. The correct classification rate was 94% for sex, 83% for age, and 61% for  
24  
25 region of origin after the leave-one-individual-out cross-validation, which is a general and  
26  
27 rigorous evaluation. The present study is expected to help guide future research to assist forensic  
28  
29 professionals to discrimination or link questioned hair and known samples from suspects or  
30  
31 victims. Future analyses could provide investigative leads to law enforcement by predicting the  
32  
33 sex, age group, and region of origin of the donor of an unknown hair sample found at a crime  
34  
35 scene.  
36  
37  
38  
39  
40

41 Profiling amino acids in human hair is just one way of examining hair. It should be  
42  
43 emphasized that this technique alone is unlikely to lead to individualization to the extent that  
44  
45 nuclear DNA can. The techniques developed through this work could complement the current  
46  
47 methods of hair analysis, which include physical examinations and genomic or mitochondrial  
48  
49 DNA analysis. The relative abundance of amino acids in hair may provide investigative leads  
50  
51 when a suspect or victim's DNA profile is not in a database. Additional studies could be carried  
52  
53 out, firstly to decrease the acid hydrolysis time and secondly to develop methods to enable  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 analysis of more than 14 amino acids, perhaps through HPLC and without the need for  
4  
5 derivatization.  
6  
7  
8  
9

### 10 **Acknowledgements**

11  
12 The Ohio University Center for Intelligent Chemical Instrumentation at Department of  
13  
14 Chemistry and Biochemistry are thanked for travel support. AHBR acknowledges financial  
15  
16 support from the Jordan University of Science and Technology. Xinyi Wang and Dr. Zhengfang  
17  
18 Wang are acknowledged for their help. GPJ acknowledges support by Grant No. 2013-DN-BX-  
19  
20 K007 awarded by the National Institute of Justice (NIJ), Office of Justice Programs (OJP), US  
21  
22 Department of Justice. The opinions, findings, and conclusions or recommendations expressed  
23  
24 in this publication/program/exhibition are those of the author(s) and do not necessarily reflect the  
25  
26 views of the Department of Justice.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## References

1. H. Sachs, *Forens. Sci. Int.*, 1997, **84**, 7-16.
2. A. S. Wilson, H. I. Dodson, R. C. Janaway, A. M. Pollard and D. J. Tobin, *Br. J. Dermatol.*, 2007 **157**(3), 450-457.
3. I. Kempson and E. Lombi, *Chem. Soc. Rev.*, 2011, **40**, 3915-3940.
4. R. Wennig, *Forens. Sci. Int.*, 2000, **107**(1-3), 5-12.
5. G. Loussouarn, C. El Rawadi and G. Genain, *Int. J. Dermatol.*, 2005, **44**(6-9).
6. K. Morioka, *Exp. Dermatol.*, 2009, **18**(7), 577-582.
7. B. Bhushan, *Biophysics of Human Hair: Biological and Medical Physics*, Springer, New York, 2010.
8. W. Rieck, *Arch. Gerontol. Geriatr.*, 1997, **25**(1), 59-71.
9. C. Robbins, *Chemical and Physical Behavior of Human Hair*, Springer, New York, 2002.
10. M. O. Han, J. A. Chun, W. H. Lee, J. W. Lee and C. Chung, *J. Cosmet. Chem.*, 2007, **58**(5), 527-534.
11. G. P. Jackson, Y. An, K. I. Konstantynova and A. H. B. Rashaid, *Sci. Justice*, DOI: 10.1016/j.scijus.2014.07.002, 2014, DOI: DOI: 10.1016/j.scijus.2014.07.002.
12. V. A. Boumba, K. S. Ziavrou and T. Vougiouklakis, *Int. J. Toxicol.*, 2006, **25**(3), 143-163.
13. M. R. Wilson, D. Polansky, J. Butler, J. A. DiZinno, J. Replogle and B. Budowle, *Biotechniques* 1995, **18**(4), 662-669.
14. C. A. Linch, S. L. Smith and J. A. Prahlow, *J. Forens. Sci.*, 1998, **43**, 305-314.
15. A. Hellmann, U. Rohleder, H. Schmitter and M. Wittig, *Int. J. Legal Med.*, 2001, **114**(4-5), 269-273.
16. J. M. Taupin, *Sci. Justice*, 2004, **44**, 95-100.
17. T. Takanohashi, *Pediatr. Int.*, 1961, **4**(1), 7-13.
18. R. C. Clay, K. Cook and J. I. Routh, *J. Am. Chem. Soc.*, 1940, **62**(10), 2709-2710.
19. A. Bertazzo, M. Biasiolo, C. V. Costa, E. Cardin de Stefani and G. Allegri, *Farmaco* 2000, **55**(8), 521-525.
20. G. Allegri, C. C., M. Biasiolo and R. Arban, *Adv. Exp. Med. Biol.*, 1991, **294**, 467-470.
21. R. H. Wilson and H. N. Lewis, *J. Biol. Chem.*, 1927, **73**, 543- 553.
22. T. A. Rutherford and P. B. Hawk, *J. Biol. Chem.*, 1907 **3**, 459-489.
23. J. Menkarr, L. Wolfram and I. Mao, *J. Soc. Cosmet. Chem.*, 1966, **17**, 769-787.
24. H. Panayiotou, Ph.D. dissertation, , Queensland University of Technology., 2004.
25. D. H. Lightbody and H. B. Lewis, *J. Biol. Chem.*, 1929, **82**, 485-497.
26. E. V. Heard and H. B. Lewis, *J. Biol. Chem.*, 1938, **123**, 203-210.
27. C. R. Robbins and C. H. Kelly, *Textile Res. J.*, 1970, **40** 891-896
28. A. H. B. Rashaid, G. P. Jackson and P. B. Harrington, *Enliven: Bio Anal Techniques*, 2014, **1**(1), 1-12.
29. R. W. Zumwalt, J. S. Absheer, F. E. Kaiser and C. W. Cherke, *J. Assoc. Off. Anal. Chem. Int.*, 1987, **70**(1), 147-151.
30. A. J. Darragh and P. J. Moughan, *J. Assoc. Off. Anal. Chem. Int.*, 2005, **88**(3), 888-893.
31. R. D. Phillips, *J. Food Sci.*, 1983, **48**, 284-285.
32. D. L. Stalling, C. W. Gehrke and R. W. Zumwalt, *Biochem. Biophys. Res. Commun.*, 1968, **31**(4), 616-622.
33. I. Molnár-Perl and Z. F. Katona, *Chromatographia*, 2000, **51**, S228-S236.

- 1
  - 2
  - 3
  - 4
  - 5
  - 6
  - 7
  - 8
  - 9
  - 10
  - 11
  - 12
  - 13
  - 14
  - 15
  - 16
  - 17
  - 18
  - 19
  - 20
  - 21
  - 22
  - 23
  - 24
  - 25
  - 26
  - 27
  - 28
  - 29
  - 30
  - 31
  - 32
  - 33
  - 34
  - 35
  - 36
  - 37
  - 38
  - 39
  - 40
  - 41
  - 42
  - 43
  - 44
  - 45
  - 46
  - 47
  - 48
  - 49
  - 50
  - 51
  - 52
  - 53
  - 54
  - 55
  - 56
  - 57
  - 58
  - 59
  - 60
34. C. W. Gehrke, H. Nakamoto and R. W. Zumwalt, *J. Chromatogr.*, 1969, **45**, 24-51.
35. C. W. Gehrke and K. Leimer, *J. Chromatogr.*, 1970, **53**(2), 201-208.
36. M. Friedman, *J. Agric. Food Chem.*, 2004, **52**, 385-406.
37. K. Robards, P. R. Haddad and P. E. Jackson, *Principles and Practice of Modern Chromatographic Methods*, Academic Press, San Diego, 1994.
38. J. Miller, *Chromatography: Concepts and Contrasts*, John Wiley & Sons Inc., Hoboken, 2005.
39. Y. An, Z. Schwartz and G. P. Jackson, *Rapid Commun. Mass Spectrom.*, 2013, **27**, 1481-1489.
40. M. P. Bartolomeo and F. Maisano, *J. Biomol. Tech.*, 2006, **17**(2), 131-137.
41. J. Gross, *Mass Spectrometry*, Springer, New York, 2004.
42. X. Shen, C. Deng, B. Wang and L. Dong, *Anal. Bioanal. Chem.*, 2006, **384**(4), 931-938.
43. C. H. Deng, X. Y. Yin, L. J. Zhang and X. M. Zhang, *Rapid Commun. Mass Spectrom.*, 2005, **19**, 2227-2234.
44. T. G. Sobolevsky, A. I. Revelsky, B. Miller, V. Oriedo, E. S. Chernetsova and I. A. Revelsky, *J. Sep. Sci.*, 2003, **26**, 1474-1478.
45. C. Vidal-madjar and G. Guiochon, *Sep. Purif. Rev.*, 1973, **2**, 1-125.
46. P. B. Harrington, *J. Chemom.*, 1991, **5**, 467-486.
47. P. D. Harrington, N. E. Vieira, P. Chen, J. Espinoza, J. K. Nien, R. Romero and A. L. Yergey, *Chemometr Intell Lab*, 2006, **82**(1-2), 283-293.
48. P. B. Harrington, C. Laurent, D. F. Levinson, P. Levitt and S. P. Markey, *Anal Chim Acta*, 2007, **599**(2), 219-231.
49. P. Rearden, P. B. Harrington, J. J. Karnes and C. E. Bunker, *Analytical Chemistry*, 2007, **79**(4), 1485-1491.
50. Z. F. Xu, C. E. Bunker and P. D. Harrington, *Appl Spectrosc*, 2010, **64**(11), 1251-1258.
51. X. B. Sun, C. M. Zimmermann, G. P. Jackson, C. E. Bunker and P. B. Harrington, *Talanta*, 2011, **83**(4), 1260-1268.
52. W. Y. Lu, J. G. Rankin, A. Bondra, C. Trader, A. Heeren and P. D. Harrington, *Forensic Science International*, 2012, **220**(1-3), 210-218.
53. F. Yang, J. Tian, Y. H. Xiang, Z. Y. Zhang and P. D. Harrington, *Cancer Epidemiol*, 2012, **36**(3), 317-323.
54. Z. F. Wang and P. D. Harrington, *Analytical and Bioanalytical Chemistry*, 2013, **405**(28), 9219-9234.
55. J. R. Wang, Z. Y. Zhang, Z. W. Zhang, Y. H. Xiang and P. D. Harrington, *Anal Methods-Uk*, 2014, **6**(19), 7695-7702.
56. J. A. Hanley and B. J. McNeil, *Radiology*, 1982, **143**, 29-36.
57. A. Gelman, *Ann. Stat.*, 2005, **33**, 1-53.
58. J. Pallant, *SPSS Survival Manual, 3rd ed.*, Open University Press, Maidenhead, 2007.
59. E. R. Malinowski, *Factor Analysis in Chemistry, 3rd ed.*, John Wiley & Sons, Inc., New York, 2002.
60. P. B. Harrington, N. E. Vieira, P. Chen, J. Espinoza, J. K. Nien, R. Romero and A. L. Yergey, *Chemom. Intell. Lab. Syst.*, 2005, **82**, 283-293.
61. D. L. Luthria, L. Z. Lin, R. J. Robbins, J. W. Finley, G. S. Banuelos and J. M. Harnly, *J. Agric. Food Chem.*, 2008, **56**(21), 9819-9827.
62. D. L. Luthria, S. Mukhopadhyay, R. J. Robbins, J. W. Finley, G. S. Banuelos and J. M. Harnly, *J. Agric. Food Chem.*, 2008, **56**(14), 5457-5462.

- 1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60
63. A. C. S. Nogueira, L. E. Dicelio and I. Joekes, *Photochem. Photobiol. Sci*, 2006, **5**, 165-169.

Table 1. Target and reference ions for SIM mode as characteristics mass fragments of the 14 amino acids.

<b>Amino acid</b>	<b>Target ion (<i>m/z</i>)</b>	<b>Reference ions (<i>m/z</i>)</b>
Ala	116	73, 147
Gly of di-TMS derivative	102	73, 147
Val	144	73, 218
Nor	144	73, 218
Leu	158	73, 102
Ile	158	73, 218
Pro	142	73, 216
Gly of tri-TMS derivative	174	73, 248
Ser	304	218, 73
Thr	73	218, 291
Asp	232	156, 73
Glu	246	73, 128
Phe	218	192, 73
Cys	220	156, 73
Lys	174	156, 73
Tyr	218	179, 280

Table 2. The FuRES leave-one-out cross-validation classification rates of sex, age group, and region of origin using each amino acid profile, averaged profiles across digestions, and averaged profiles across individuals.

Factor	Correct classification rate		
	Profiles ( <i>n</i> =528)	Average Across Trials ( <i>n</i> =177)	Average Across Individuals ( <i>n</i> =64)
Sex	84%	85%	94%
Age group	88%	80%	83%
Region of origin	81%	68%	61%

Table 3. ANOVA results of the effect of sex on the mean squares of variance between and within groups of the concentrations ( $\mu\text{mol/mL}$ ) of amino acids.

Amino Acid	Male Mean $\pm$ 95% CI ( <i>n</i> =33)	Female Mean $\pm$ 95% CI ( <i>n</i> =31)	Mean squares of Variance (between groups)	Mean squares of variance (within groups)	<i>F</i> -value	<i>p</i> - value
<b>Ala</b>	0.24 $\pm$ 0.04	0.21 $\pm$ 0.05	0.005	0.03	0.02	0.6
<b>Gly</b>	0.08 $\pm$ 0.03	0.07 $\pm$ 0.02	0.006	0.002	0.2	0.4
<b>Val<sup>e</sup></b>	0.09 $\pm$ 0.01	0.13 $\pm$ 0.03	0.3	0.004	7	0.004
<b>Leu<sup>e</sup></b>	0.18 $\pm$ 0.04	0.17 $\pm$ 0.02	0.01	0.01	0.9	0.2
<b>Ile<sup>e</sup></b>	0.054 $\pm$ 0.008	0.051 $\pm$ 0.001	0.001	0.001	2	0.05
<b>Pro</b>	0.14 $\pm$ 0.02	0.18 $\pm$ 0.01	0.2	0.006	9	0.02
<b>Ser</b>	0.04 $\pm$ 0.01	0.03 $\pm$ 0.09	0.001	0.001	0.005	0.9
<b>Thr<sup>e</sup></b>	0.036 $\pm$ 0.009	0.022 $\pm$ 0.007	0.04	0.007	11	0.005
<b>Asp</b>	0.12 $\pm$ 0.02	0.09 $\pm$ 0.03	0.05	0.01	0.9	0.2
<b>Cys</b>	0.04 $\pm$ 0.01	0.03 $\pm$ 0.01	0.002	0.002	0.1	0.3
<b>Glu</b>	0.08 $\pm$ 0.02	0.09 $\pm$ 0.02	0.1	0.003	7	0.04
<b>Phe<sup>e</sup></b>	0.211 $\pm$ 0.002	0.321 $\pm$ 0.001	0.2	0.003	13	0.001
<b>Lys<sup>e</sup></b>	0.082 $\pm$ 0.001	0.081 $\pm$ 0.001	0.004	0.001	0.09	0.8
<b>Tyr</b>	0.051 $\pm$ 0.006	0.048 $\pm$ 0.005	0.002	0.001	0.3	0.3

<sup>e</sup>Essential amino acid

Table 4. ANOVA results of the effect of age group on the mean squares of variance between and within groups.

Amino Acid	Group A Mean $\pm$ 95% CI ( <i>n</i> =54)	Group B Mean $\pm$ 95% CI ( <i>n</i> =10)	Mean squares of variance (between groups)	Mean squares of variance (within groups)	<i>F</i> -value	<i>p</i> - value
Ala	0.22 $\pm$ 0.03	0.14 $\pm$ 0.01	0.6	0.02	10	0.005
Gly	0.05 $\pm$ 0.04	0.08 $\pm$ 0.03	0.7	0.002	23	0.004
Val <sup>e</sup>	0.14 $\pm$ 0.01	0.11 $\pm$ 0.08	0.004	0.005	0.8	0.4
Leu <sup>e</sup>	0.18 $\pm$ 0.01	0.17 $\pm$ 0.01	0.03	0.01	0.3	0.1
Ile <sup>e</sup>	0.06 $\pm$ 0.06	0.03 $\pm$ 0.03	0.03	0.001	5	0.03
Pro	0.14 $\pm$ 0.01	0.15 $\pm$ 0.01	0.006	0.006	1	0.3
Ser	0.025 $\pm$ 0.008	0.03 $\pm$ 0.02	0.002	0.001	0.1	0.5
Thr <sup>e</sup>	0.034 $\pm$ 0.006	0.025 $\pm$ 0.002	0.07	0.007	8	0.001
Asp	0.11 $\pm$ 0.02	0.08 $\pm$ 0.05	0.05	0.01	6	0.02
Cys	0.031 $\pm$ 0.009	0.04 $\pm$ 0.04	0.02	0.008	3	0.01
Glu	0.08 $\pm$ 0.01	0.09 $\pm$ 0.03	0.001	0.003	0.4	0.07
Phe <sup>e</sup>	0.04 $\pm$ 0.02	0.023 $\pm$ 0.002	0.001	0.002	1	0.06
Lys <sup>e</sup>	0.078 $\pm$ 0.008	0.08 $\pm$ 0.04	0.02	0.002	0.8	0.4
Tyr	0.048 $\pm$ 0.004	0.05 $\pm$ 0.02	0.001	0.01	0.3	0.6

<sup>e</sup>Essential amino acid

Table 5. ANOVA results of the effect of the region of origin on the mean squares of variance between and within group of the abundance of 14 amino acids.

Amino Acid	Mountain Mean $\pm$ 95% CI (n=29)	Valley Mean $\pm$ 95% CI (n=23)	Desert Mean $\pm$ 95% CI (n=12)	Mean squares of variance (between groups)	Mean squares of variance (within groups)	F-value	p-value
Ala	0.19 $\pm$ 0.01	0.24 $\pm$ 0.03	0.19 $\pm$ 0.02	0.3	0.02	6	0.001
Gly	0.08 $\pm$ 0.05	0.08 $\pm$ 0.08	0.081 $\pm$ 0.008	0.006	0.02	0.4	0.4
Val <sup>e</sup>	0.105 $\pm$ 0.009	0.11 $\pm$ 0.01	0.106 $\pm$ 0.008	0.002	0.05	0.3	0.1
Leu <sup>e</sup>	0.19 $\pm$ 0.01	0.18 $\pm$ 0.02	0.17 $\pm$ 0.01	0.03	0.1	2	0.05
Ile <sup>e</sup>	0.05 $\pm$ 0.03	0.06 $\pm$ 0.04	0.05 $\pm$ 0.05	0.1	0.001	20	0.001
Pro	0.15 $\pm$ 0.09	0.149 $\pm$ 0.001	0.138 $\pm$ 0.001	0.03	0.006	5	0.004
Ser	0.03 $\pm$ 0.05	0.03 $\pm$ 0.04	0.03 $\pm$ 0.04	0.007	0.001	0.5	0.2
Thr <sup>e</sup>	0.03 $\pm$ 0.03	0.03 $\pm$ 0.03	0.03 $\pm$ 0.03	0.05	0.001	8	0.6
Asp	0.09 $\pm$ 0.01	0.11 $\pm$ 0.02	0.10 $\pm$ 0.05	0.05	0.009	6	0.004
Cys	0.021 $\pm$ 0.007	0.034 $\pm$ 0.004	0.054 $\pm$ 0.003	0.4	0.002	24	0.001
Glu	0.08 $\pm$ 0.06	0.09 $\pm$ 0.08	0.08 $\pm$ 0.08	0.003	0.003	0.5	0.8
Phe <sup>e</sup>	0.028 $\pm$ 0.001	0.026 $\pm$ 0.006	0.021 $\pm$ 0.002	0.01	0.002	18	0.001
Lys <sup>e</sup>	0.08 $\pm$ 0.05	0.09 $\pm$ 0.06	0.08 $\pm$ 0.04	0.02	0.004	7	0.04
Tyr	0.054 $\pm$ 0.003	0.047 $\pm$ 0.003	0.0347 $\pm$ 0.002	0.04	0.0001	12	0.002

<sup>e</sup>Essential amino acids

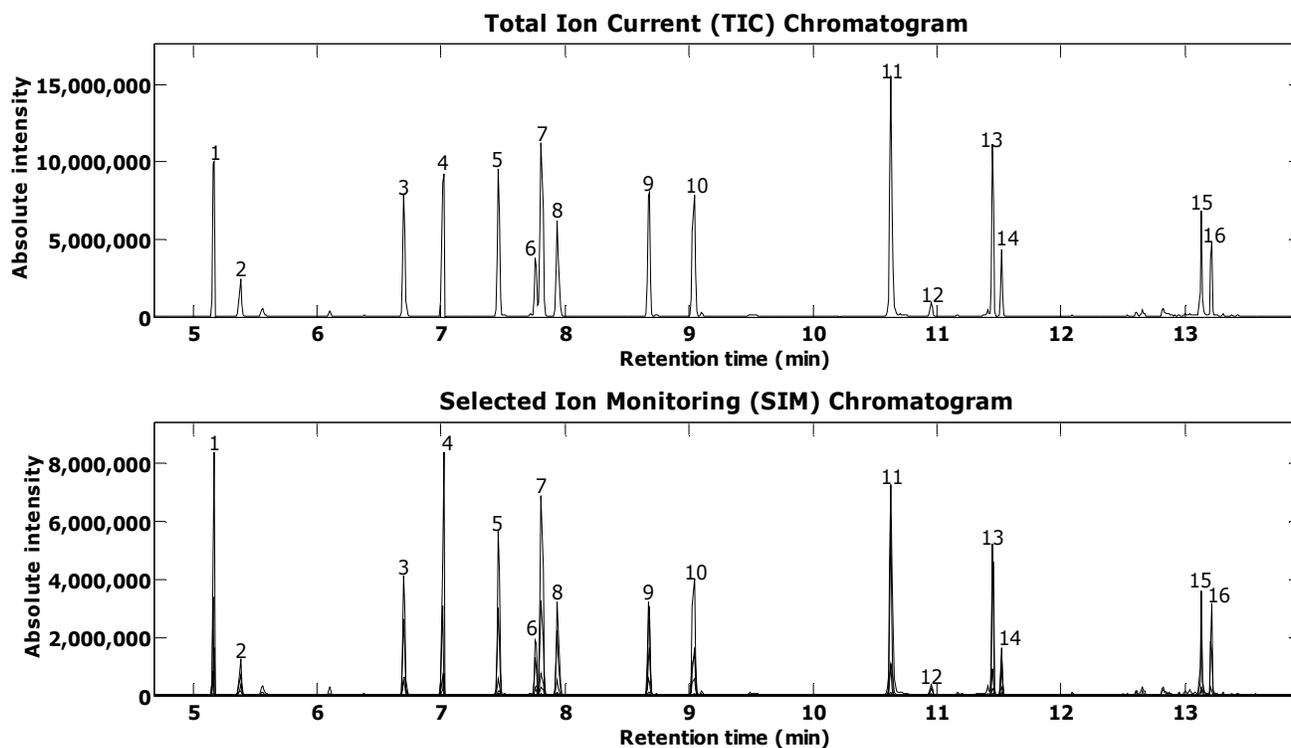


Fig. 1 A typical GC/MS profile for the analysis of human scalp hair keratin protein after derivatization with BSTFA at 100 °C for 30 min. SIM shows signals for target and reference ions that were described in GC/MS analysis step of the experimental section. Peaks: (1) Ala; (2) Gly2\*; (3) Val; (4) Nor; (5) Leu; (6) Ile; (7) Pro; (8) Gly3\*\*; (9) Ser; (10) Thr; (11) Asp; (12) Cys; (13) Phe; (14) Glu; (15) Lys; (16) Tyr.

\*Di-trimethylsilyl Derivative

\*\* Tri-trimethylsilyl Derivative

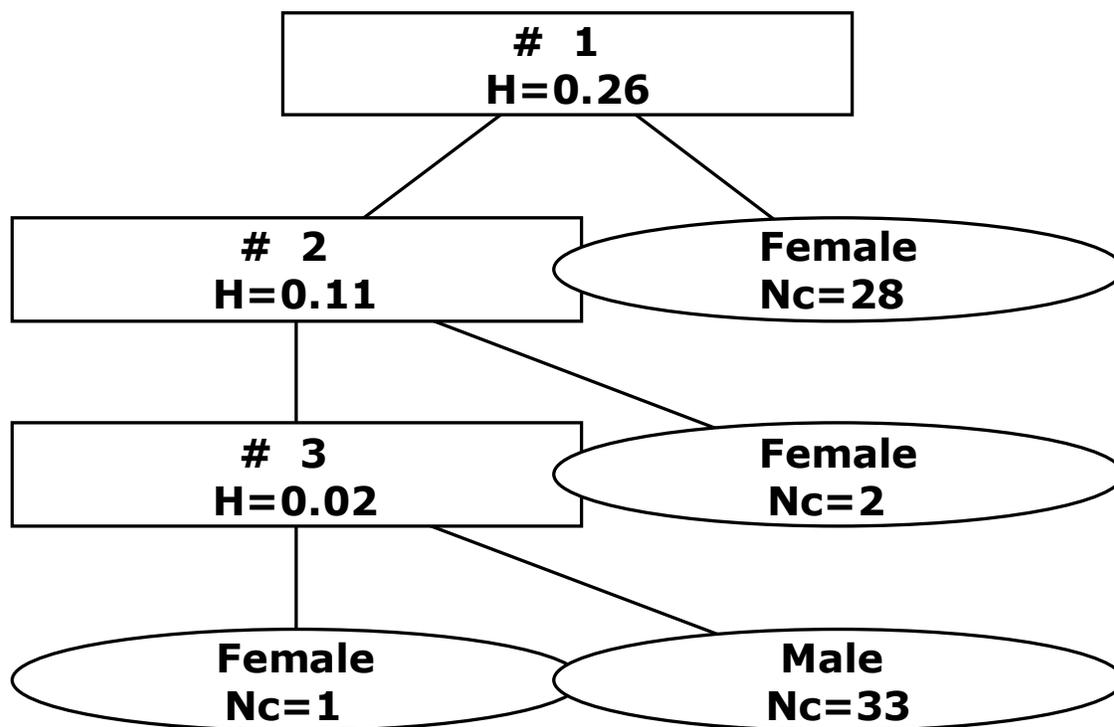


Fig. 2 FuRES sex classification tree from the average profiles of 14 amino acids in the hair for 64 individuals. This tree yielded a 94% prediction rate. Nc is number of individuals sorted to each node by the rules in the tree. H is the fuzzy entropy of classification for each rule and the rule number is given in each rectangle. The circular nodes are terminal and designate the class.

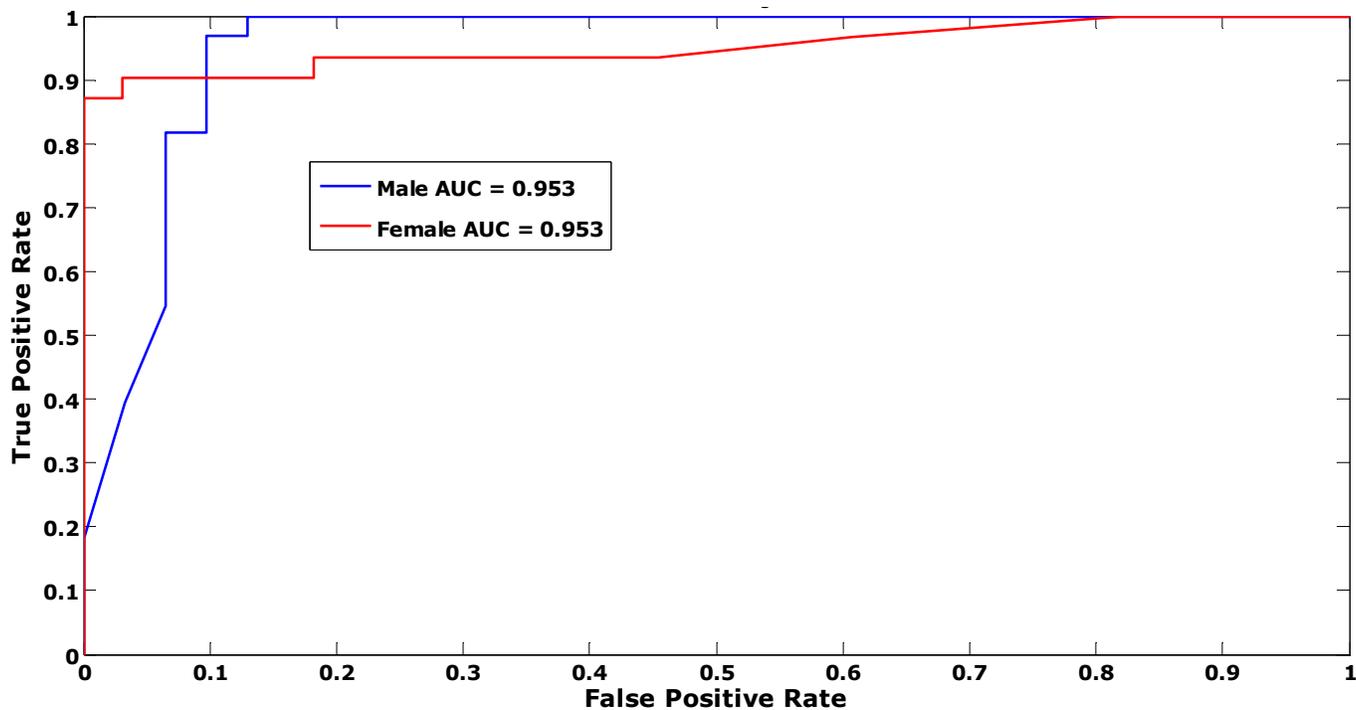


Fig. 3 Receiver operating characteristic (ROC) curve for identifying the sex of 64 subjects.

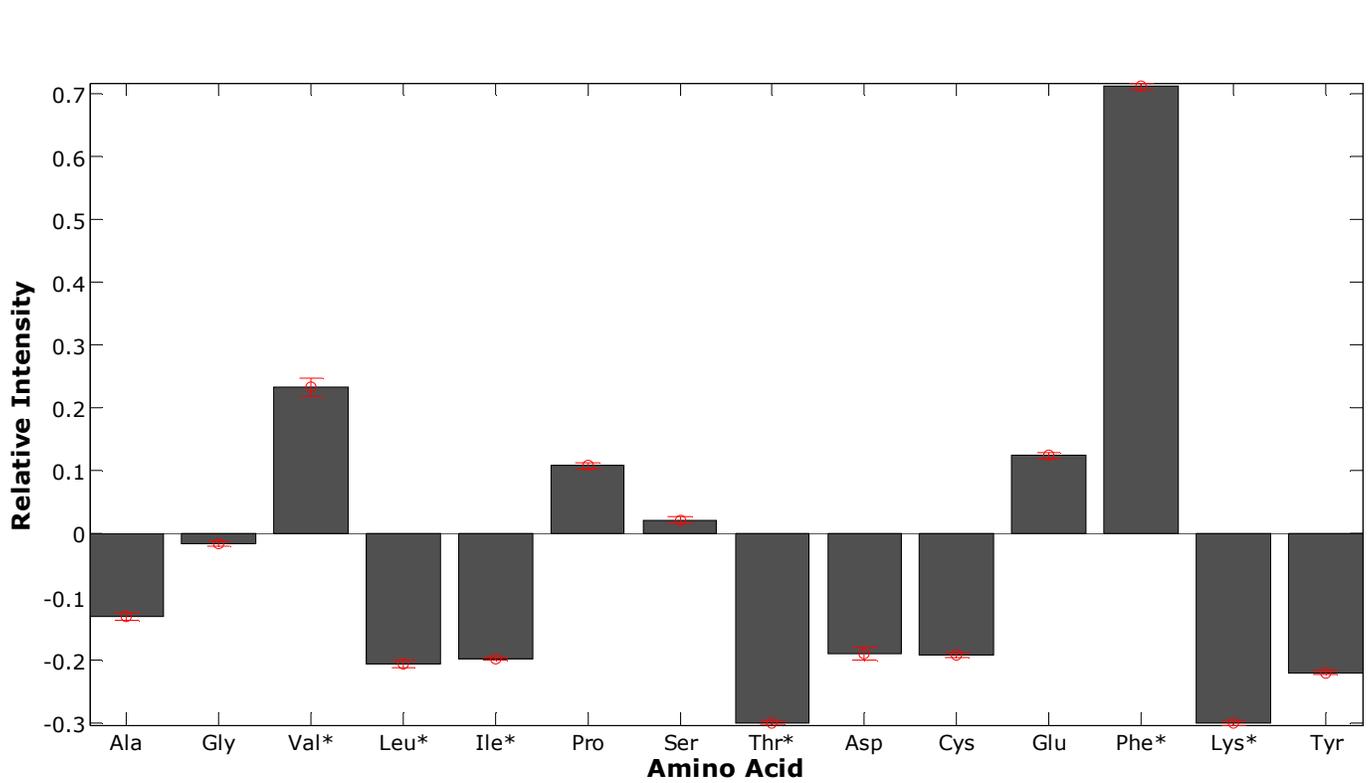


Fig. 4 Variable loadings scores of the 14 amino acids of sex classification.

\*Essential amino acids.

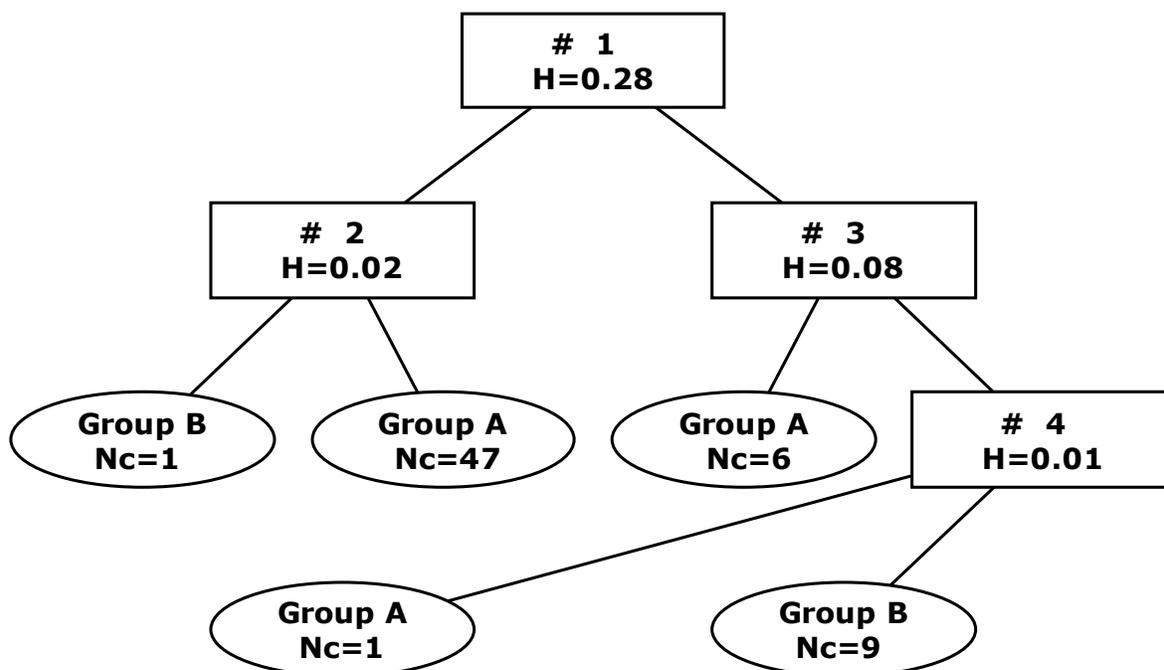


Fig. 5 FuRES age group classification tree from the average amino acid profiles of 14 amino acids in the hair from 64 subjects (n of 9 for each average) with an 83% classification rate. H is the entropy of the classification rule and the number of rules required to build the tree are given in the rectangles. The number of subjects and name of the group are written in each leaf node.

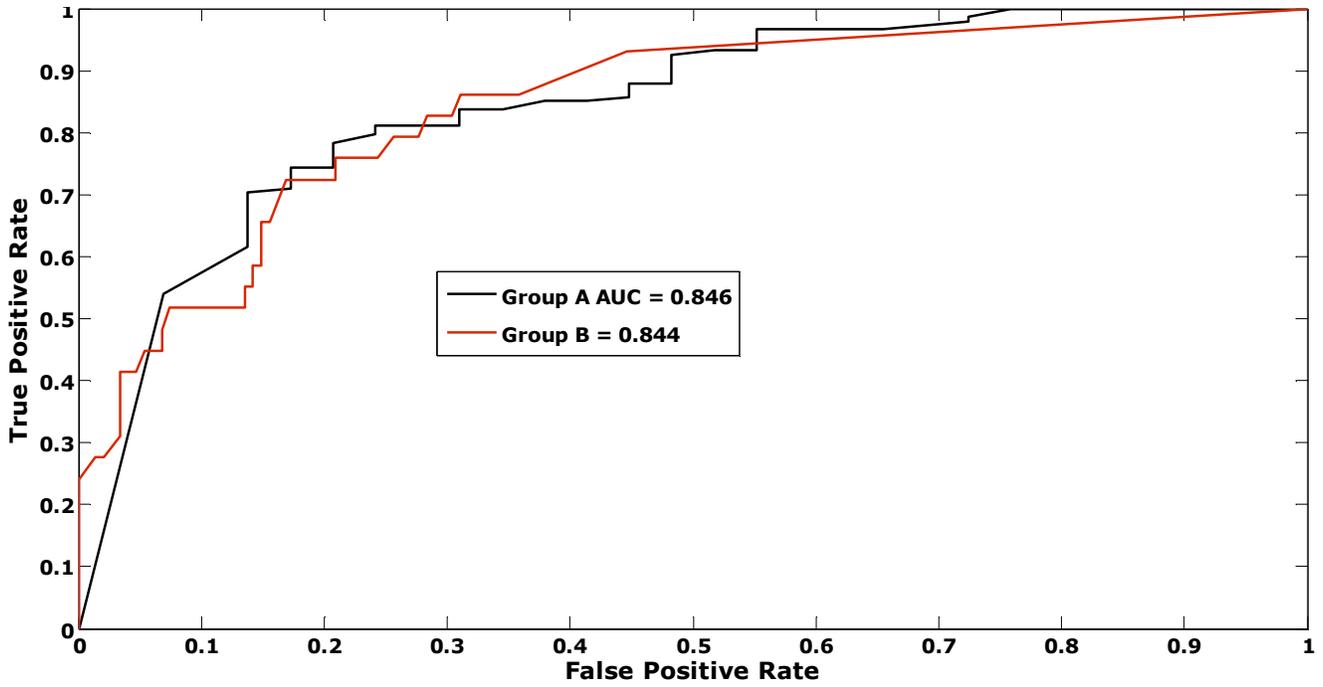


Fig. 6 Receiver operating characteristic (ROC) curve for identifying age group (1-49 yrs or 50-77 yrs) of the 64 subjects.

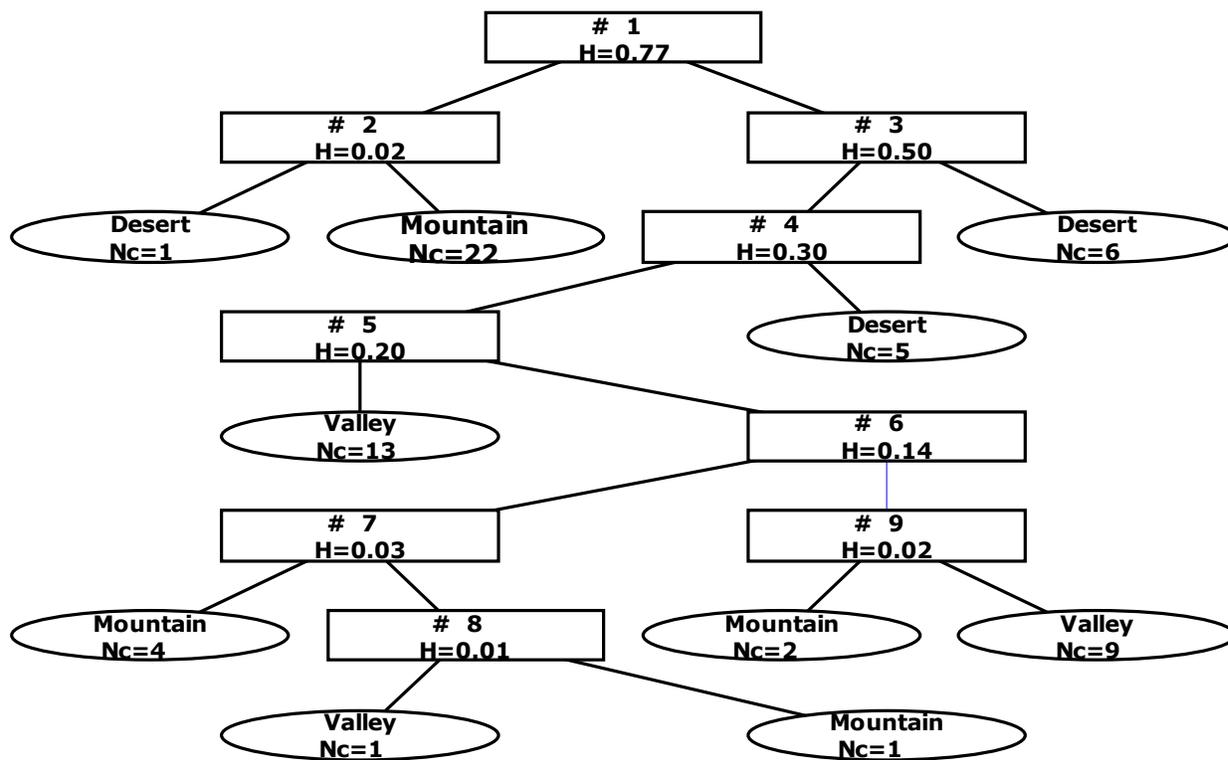


Fig. 7 FuRES region of origin classification tree from the relative abundance of 14 amino acids in the hair of 64 subjects ( $n$  of 9 for each subject) with a 61% classification rate.

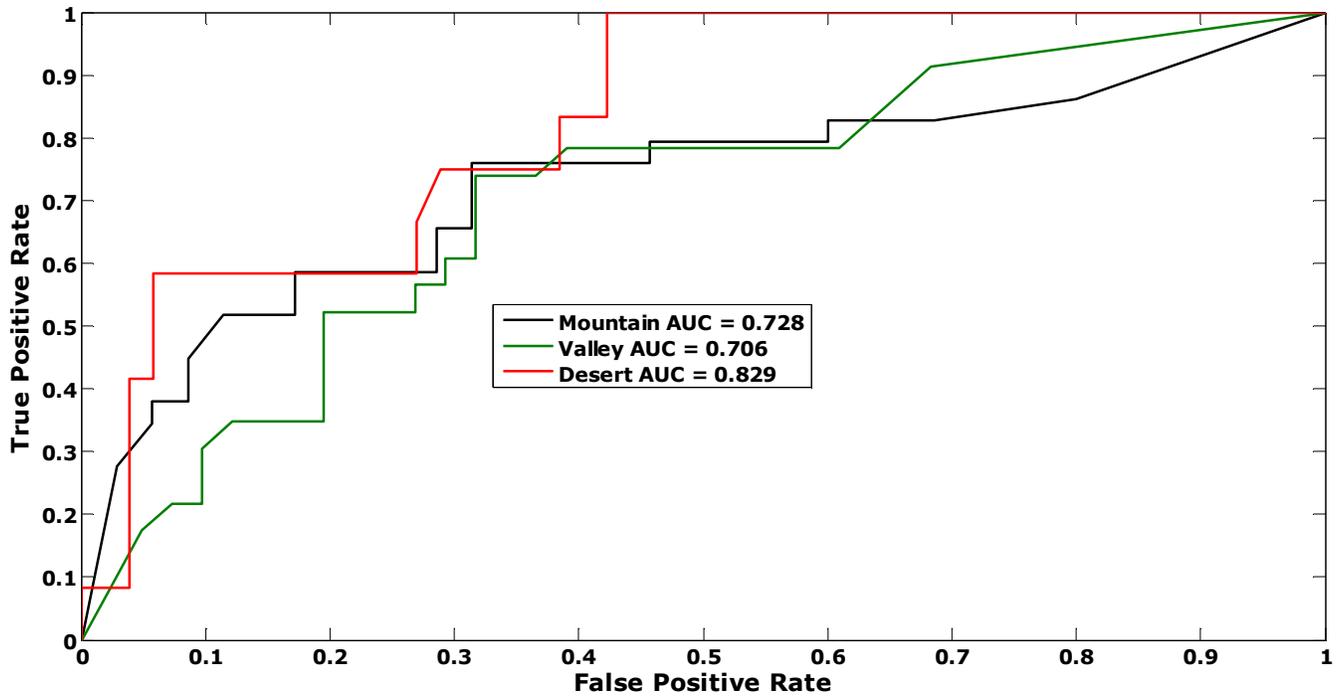


Fig. 8 Receiver operating characteristic (ROC) curve for identifying the region of origin of 64 subjects.