Historical Immigrations to Shanghai Suburb Based on Principal Component Analysis on Physical characters and Surnames

Hui Li[1], Liqun Xu [2], Zhenhe Zhou[3], Lingjun Zhang[4], Jianzhong Jin[1],
Daru Lu[1], Li Jin[1*]

1. Center for Anthropological Studies, School of Life Sciences, Fudan University, Shanghai 200433,China.
2. Department of Statistics, Management College, Fudan University, Shanghai 200433,China.
3. Center for Historical Geographic Studies, Fudan University, Shanghai 200433,China.
4. Department of Biotechnology, Shanghai Fisheries University, Shanghai 200090,China.

Principal component analysis based on physical characters and surnames divides Shanghai suburbanites from 13 townships of 9 counties into 3 clusters. From the result of other research fields, the 3 clusters shall be Yue (Daic), Wu and historical immigrants from northern China. The variance percentages of the components show that the primary difference in the suburbanites of Shanghai is between northern Immigrants and southern Natives. Secondary difference is between two kinds of Southern Native, Wu and Yue, and then it divides Yue into two parts with different origins. Changing the components to the contour on the Shanghai maps, the historical immigration events are shown in different maps. Geographic analysis on surname components also tells that the diffusion features and patterns of southern surnames and northern surname are entirely different. Our study not only answers the question where Shanghai suburbanites come from, but also provides a new approach for immigration history research. In the following researches, more DNA sequencing and HLA typing will be applied.

Key words: Shanghai suburbanites, Historical immigrations, Principal component analysis, Wu-Yue.

Introduction

As a cosmopolis, Shanghai history is short for only about a century; but it has a long history of people's settlement. The discoveries of Archaeology emerge incessantly to prove that

there are many people settling here over five or six thousand years ago [1]. Along with the years, many occurrences brought the immigrants group by group to Shanghai, and the immigrants became the important study objects of archaeology research. The special location of Shanghai endows the suburbanites with a particular status among the Chinese populations. The Genetic study proves that, Yangtze River is the approximate division of North and South Chinese populations [2]; and Shanghai is located on one end of this boundary, influenced by both North and South Chinese populations, so that the study of its suburbanites has a profound significance to the study of genetic structure of Chinese populations. In the history, Shanghai once was located along the border of Wu Kingdom and Yue Kingdom, and is the borderline of Jiangsu Province and Zhejiang Province all the while, so the historical immigrations involve many historical events. Furthermore, when the study of anthropology involved in China, there were a few scholars focusing their sight on Shanghai suburbanites.

In the middle of the 19th century, some occidental missionaries pointed out in their literatures that the populations of shanghai used different dialects and had different dressing cultures [3]. Then the recognition to Shanghai suburbanites did not improve much during a long time. In beginning of the 20th century (1928), Yuanren Zhao pointed out that according to his study of Wu dialect, in the Wu dialect area (including Shanghai) today primarily lived people of Yue nation during Spring and Autumn Period. And along with the continuous emigrant action of Yue's moving northward and Han's moving southward since Qin and Han Dynasty, formed the pattern of Yue and Han poly-living [4]. This is a breakthrough of the recognition to Shanghai suburbanites. Since the 80s of the 20th century, the studies of linguistics and anthropology have improved quiet greatly and many new standpoints have been emerged. In 1987, Nairong Qian found that the difference in dialects between eastern and western of Fengxian county of Shanghai is great when he studied them [5]. In 1988, Shangfang Zhengzhang analyzed some dialects' special speeches of south Zhejiang province and Shanghai, conferred that the ancient dwellers of the two regions not only were both Yue, but may had closer relationship, and differed from Wu, so that he believed that there should be quite a few descendants of Yue among the present dwellers [6]. In 1993, Baohua Xu, etc. finished distributing Shanghai with dialects [7]; the people of the seven regions may have different origins. At that time, the study of Shanghai suburbanites is limited in linguistics and cultural anthropology.

In 2000, our research team began the physical anthropology and genetics studies on the Shanghai suburbanites. We first found that the populations between eastern and western Fengxian are different according to their physical characters as well as their dialects [8]. The genetic analysis of Maqiao region and western Fengxian shows that the Y chromosome of the suburbanites of this coterminous region is not the normal constitution of Han Chinese, but the same to other Daic (Yue) populations, a few chromosomes can even trace back to some present regions such as Laos [1]. It also proves the reference of linguistics that there are a few descendants of Yue among the present dwellers. After that, we researched for the physical characters and surnames of the populations of the 13 townships locating at Shanghai suburb, and found that according to the descent of physical characters and surnames, Shanghai suburbanites are divided into 3 clusters, namely the descendants of Yue, Wu and north immigrates [9,10]. Through the analysis of intermarriage mechanism, we pointed out that the change of the Shanghai populations is not aroused by intermarriage, but primarily by migration. Because of the different origin, the three clusters live separately, without obvious mixing [10]. Based on these researches, we went deep into analyzing what migration occurrences had left traces among the Shanghai suburbanites and the historical occurrences corresponding to the traces, in order to make a conclusion of our research. Utilizing the published relative data of the 13 populations

in the paper, with the statistical method of principal component analysis [2,11], we drew the tendency of the difference between populations on the map we established.

Materials and Methods

Objects

We researched the Shanghai ethnic suburbanites from 13 townships (Jinhui, Touqiao, Shangta, Zhujiajiao, Tianmasha, Shiwan, Shuyuan, Fengjing, Langxia, Puzhen, Nanxiang, Zhujiaqiao, Maqiao) of nine counties, and their particular geographical locations are mapped on Fig.1. On this figure, the most population characteristic headscarves, which illustrate the cultural difference between populations can also be seen. The ecumenical status of investigated objects has been depicted at relative papers [9,10].
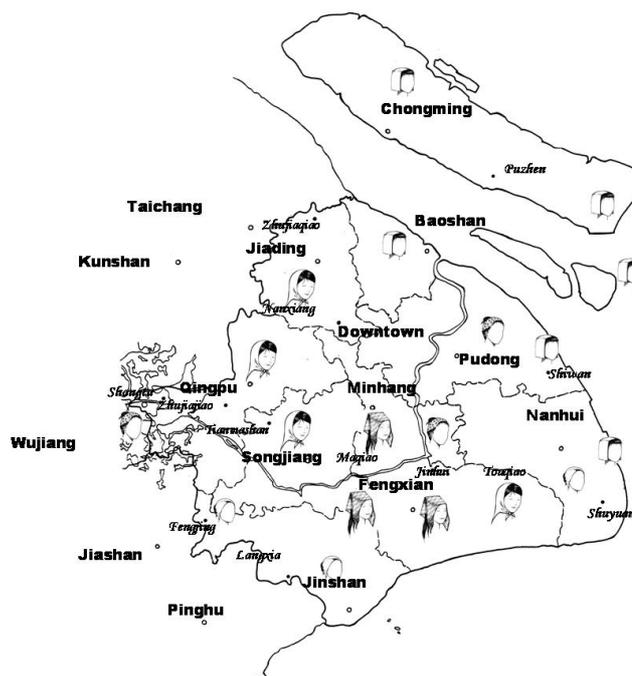

Fig.1.   Location and culture(shown with headscarves such as babushka, coif and kerchief) of the populations we studied

Item Analysis

We have listed the observational facial features, metrical head and body data and standardized index, dermatoglyphic parameters of the 13 populations[9]. With different heritabilities of every parameter to the populations, and the analysis demand on standard data, we select all the observational face features and standard metrical physical parameters of qualitative significance to make further analysis. The surnames of the 13 populations have been listed previously [10].

Principal Component Analysis

We use the SPSS (Statistical Product and Service Solutions) 10.0, which is a statistical analytic software developed by SPSS Co. Ltd. in USA to do the principal component analysis[11]. We used the first and second principal components to construct the rectangular coordinates and drew the diffused–dot plot of every population, in order to observe the difference between them. According to the geographical distribution of populations, we drew the grade distribution plots of every principal component onto Shanghai map with the principium of drawing contour lines [2]. We used the data of surnames to do principal component analysis, drew the diffused-dot plot previous, and made it to draw the grade distribution plots in this paper.

Results and Discussion

Principal component analysis of the relevant physical character data.

We make principal component analysis to the selected physical character data, and have received 13 principal components. The first to tenth component possess the main variance, therefore they are more significant. You can see the variance percentage of them in Tab.1.

Tab.1 Variance percentages of components from principal component analysis of physical characters of 13 populations in the suburb in Shanghai

Principal Component	1	2	3	4	5	6	7	8	9	10
Variance Percentage(%)	84.17	9.25	2.01	1.07	0.93	0.60	0.53	0.38	0.35	0.27

Thus, the first and second principal components include the majority of the data information, representing the two main differences among Shanghai suburbanites, and the suburbanites can be divided into 3 origins. They correspond with the previous inference that Shanghai suburbanites can be divided into 3 origins, namely the descendants of Yue, Wu and north immigrates. The rest include the faint trace left by other historical migration, too.

Three origins of Shanghai suburbanites

We use the first and second principal components of every population to construct the rectangular coordinates; it is seen that the coordinate distributing of every population on Fig.2.
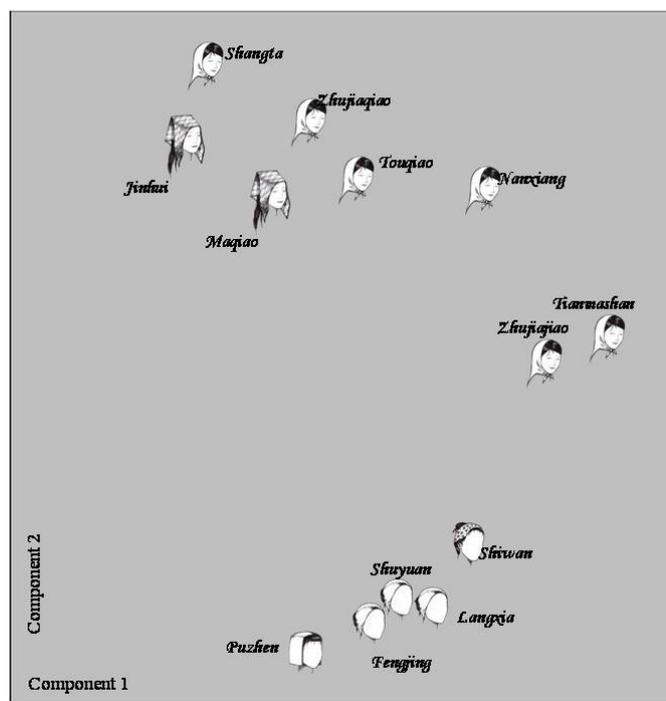

Fig.2 Distances among populations shown by principal component analysis of physical characters

On the figure it is seen that the populations with the same or similar culture characters gather closely, which illuminates the relative close distance in their physical characters. Therefore, the 13 populations can be divided into 3 clusters. Jinhui and Maqiao have the same culture and similar physical characters, and we have proved that they would be most likely the Yue's descendants [1]. To depict briefly, we directly call them Yue in the following. Chongming, Pudong, Jinshan represented respectively by Puzhen, Fengjing, Langxia, Shiwan, Shuyuan are comparatively close, their culture features are similar to those of the people of Wujiang, Wuxian, etc; Maybe they originate from Wu; we directly call them Wu in the following. The rest populations have close culture, but their physical

characters are comparatively dispersed, maybe they are from relatively tangled originated northern immigrants. Although the three clusters affect each other, they keep their relative independent, and can be distinguished. We analyze the geography with the principal component in the following, and research on the process of the three clusters to form the present distribution pattern.

The geography analysis of physical characters

From frg.3 to fig.12, we draw the geographical distribution of every principal component, with each distribution representing one of the migration occurrences. We analyze and explain the first to third principal component distributions.
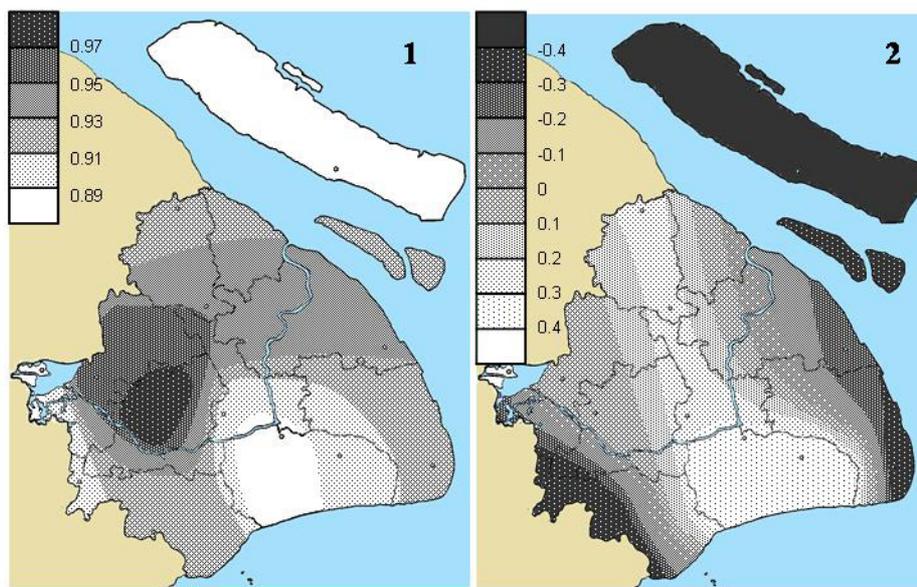

Fig.3 Geographic analysis of the first component of physical characters. Fig.4 Geographic analysis of the 2nd component of physical characters.

The first principal component is a typical tendency of diffusion from Songjiang to surrounding districts. Songjiang is the first step of feudal officials in Shanghai history, with the central administration's setting up the political institution, certainly many Chinese Centrality immigrants have come here. The predomination of the feudal official engendered a series of guarantees on politics, economics, culture, military affaires, which attract the northern immigrants to settle down here. In general, it is hard for northern immigrants to live in the area of southern aborigines, so the diffusion on the map detours the Yue's area such as Maoqiao. Previously, Songjiang stood for the political center of Shanghai for the longest time, so that this tendency took the biggest part of all. This principal component can be regarded as not only the diffusion and settlement of Songjiang district people, but also the difference between northern immigrants and south aborigines, Wu and Yue. Therefore, the greatest difference among Shanghai suburban populations is the one between north and south.

The second principal component is a typical tendency of populations blending like grades with one kind of populations standing in the middle, another at two sides. The populations in the middle live at the Gangshen shell dyke running from northwest to southeast through Shanghai district. Gangshen shell dyke is a series of shell sandy bank formed about four thousand years ago, the averagely highest place in physiognomy of Shanghai, and the belt people began to live earlier. According to the archaeological discoveries and historical descriptions, it is three thousand years ago when the Wu's Hushu culture came to the Tai Lake Plain [12]. Therefore, most of the dwellers of this

belt should be Yue, with distribution of more in south and less in north, which inosculates the distribution of culture. In addition, the people of two sides are Wu, with a slightly gradual transition from Yue. And the transition is steep in south and tempered in north, because northern immigrants impact the north. Thus, it can be seen that the secondary difference than the one between north and south is the one between two kinds of southern aborigines, Wu and Yue.
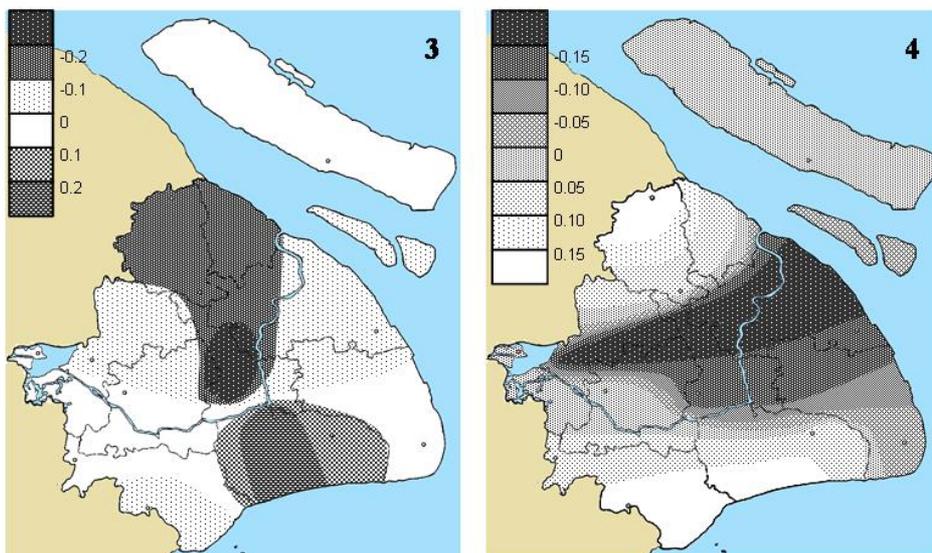

Fig.5 Geographic analysis of the 3rd component of physical characters. Fig.6 Geographic analysis of the 4th component of physical characters.

The third principal component is a tendency of two populations' blending, and is different between Maqiao and Jinhui. This shows that although the culture of Yue is consistent, the populations may be from two origins. Some of the Yue people of Eastern Ouyue Kingdom (Dong-Ou) at southern Zhejiang migrated to Jiang-Huai (area along Yangtze River) in Han Dynasty, but there is no saying about where they have migrated. It is very possible that they came to Shanghai, and neighbored with the Yue coming earlier in Shanghai, because the populations of the same nation go along with each other comparatively well. It is hard to decide whether Maqiao or Jinhui is the descendant of Eastern Ouyue Kingdom. The archaeological cultures between northern and southern Zhejiang province are not consistent[11], so that the differences of *Yue* between Maqiao and Jinhui are reasonable. Therefore, the third difference between Shanghai suburbanites is the interior difference of Yue.

The fourth principal component and the rest can be idiographic meaning, but they involve little data, and the significance of them is undefined.
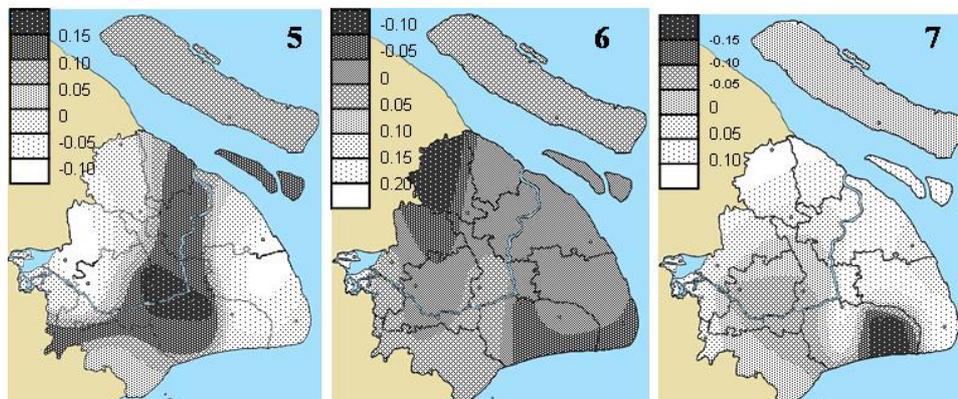

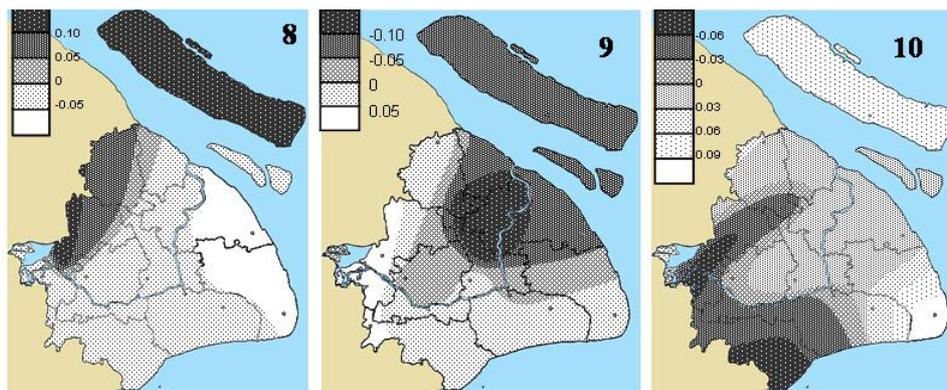Fig.7,8,9　Geographic analysis of the 5th, 6th and 7th components of physical characters.

Fig.10,11,12   Geographic analysis of the 8th, 9th and 10th components of physical characters.

Result of the surnames' principal component and the geographical analysis

The research of the structure of the 13 populations' surnames can also give clue to the origin of every family. The principal component of surnames also separates out 13 principal components, variance percentages of which are much closer than the result of physical character analysis. Because surnames are not the naturally genetic feature, and cannot reflect many phenomenon like physical characters, we only observe 1st to 4th principal components, and list the variance percentages of the date on Table 2.

Tab.2 Variance percentages of components from principal component analysis of surnames of 13 populations in the suburb of Shanghai

Principal component	1	2	3	4
Variance Percentage(%)	51.35	8.34	6.87	6.35

The principal component diffused-dot plot of the population has been drawn in another paper [10]. Because Yue people had no Chinese surnames before, the present surnames referred to the ones of neighbor populations in history. Therefore, there is no Yue population separated out on the diffused-dot plot.

From Fig13 to Fig 16 we show the geographical grade distribution of 1st to 4th principal components.

The first principal component diffused from the Suzhou district in west to all of Shanghai, and the diffusion is a single debasement of the principal component. It should represent the diffusion of the south families and their surnames. Because the south families and surnames are more in population and less than the north immigrants, their families' percentage in suburbanites diminishes gradually from west to three other directions, but the kinds of surnames are not finished. It is the biggest origin of the suburban families.

The second principal component shows two different tendencies from middle and north of Shanghai to southwest and southeast. It is delegate the diffusion of the north families and surnames. The north immigrants group by group brought many kinds of surnames, and have the surname characters of many kinds and few people in each surname. As a result, the immigrants carried different surnames after they came from north and went to two different directions, showing the movement of the principal component from zero to positive and negative sides.
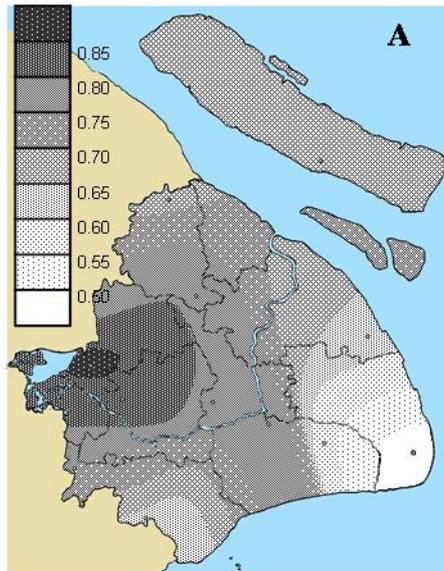
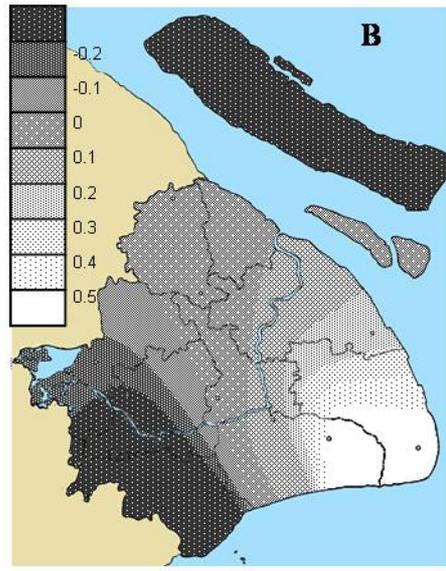Fig.13.   Geographic analysis of the 1st component of surnames.   Fig.14.   Geographic analysis of the 2nd component of surnames.
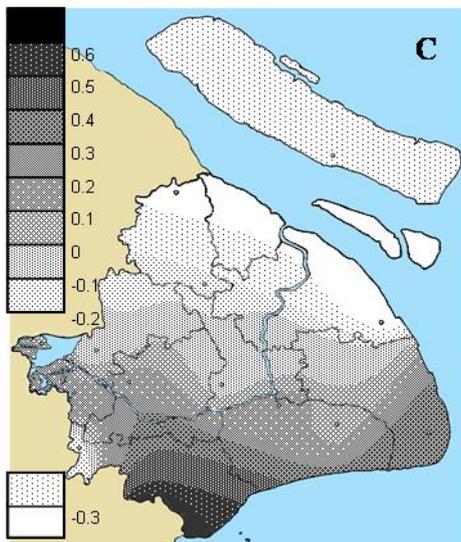

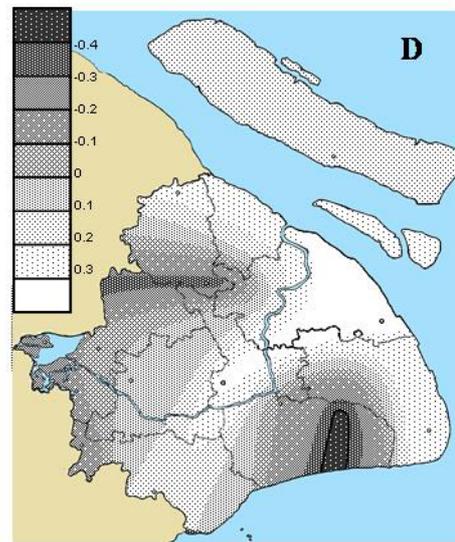Fig.15.   Geographic analysis of the 3rd component of surnames.   Fig.16.   Geographic analysis of the 4th component of surnames.

The third principal component shows the geographical inter-transition tendency of the families from Zhejiang and Jiangsu. Their division (zero degree line) leans to Zhejiang comparatively to the ancient boundary between Zhejiang and Jiangsu. It illustrates the immigrants from Jiangsu were much more than from Zhejiang.

The fourth principal component may be the phenomenon that the western paludal belt and eastern seashore reclaiming and cultivating families came in.

The significance for the research of migrant history

The principal component analysis on physical characters and surnames of Shanghai suburbanites does not only plot out the 3 origins of Wu, Yue and north immigrants, but also outline the many emigration occurrence that had left traces among Shanghai suburbanites. It should greatly enlighten the study of migration history. Since we can analyze out so much phenomenon at Shanghai of unusual complexity, we can discover the migration traces more clearly at other places, and resolve some suspenseful questions in many literature.

The variance percentages of every principal component occupying the data are also very important,

and reflect the proportion of immigrants at a certain extent. For example, the principal component analysis of physical characters in this paper figures out the sum of principal component (2,3,5) variance involving Yue is about 12.2%, which approximately equal to the proportion of 400,000 people living in Maoqiao, etc. with Yue custom among the Shanghai suburbanites.

We believe that the study form using mathematical statistical method combined with physical anthropological research will certainly promote the migration history study.

References

1. Li H: A Genetic Study on Genes of Human Bones in Graves. In: The CPAM of Shanghai. Maqiao report on Excavation between 1993 and 1997. Shanghai: Shanghai Fine Arts Publisher,2002;57-68.

2. Xiao CJ, Cavalli-Sforza LL, Minch E, Du RF: Principal Component Analysis of Gene Frequencies of Chinese Populations. Sci China Ser C, 2000;43(5): 472-481.

3. Zhou TC: Shanghai Speech in Nineteenth Century. In: Wu-Dialect Research Studio, Institute of Chinese Language and Literature. Collected Papers on Wu-Dialect. Shanghai: Shanghai Education Press, 1988;175-183.

4. Qian NR: Studies on Modern Wu-Dialect. Shanghai: Shanghai Education Press,1992;2.

5. Qian NR: Linguistic Binary Line between East and West Fengxian County. In: Li ZL: The First Collection on Linguistic Studies Shanghai: Press of Fudan University,1987;297-308.

6. Zheng-Zhang SF: The Tight-throated Plosive Sonant in Dialects in South Zhejiang and Shanghai[A]. In: Wu-Dialect Research Studio, Institute of Chinese Language and Literature. Collected Papers on Wu-Dialect. Shanghai: Shanghai Education Press, 1988;232-237.

7. Xü BH, Tang ZZ, Chen ZM. Classification and Distribution of Dialects in Shanghai. *Dialects*.1993; 1:14-30.

8. Lin L, Li H, Chen L, Sun PH, Xie KQ, Xia YM, Qiao SY, Lu DR, Jin L: Physical characters Discrepancy Between Two Populations in East and West Fengxian County. Journal of Fudan University(Natural Science),2002;41(1):97-101.

9. Lin L, Li H, Zhang HG, Xia YM, Jin L, Lu DR: Physical constitution and genetic relationship of Shanghai Suburbia. Acta Anthropologica Sinica,2002;21(4) :293-306.

10. Xu LQ, Li H, Xi HF, Jin L: The Analysis of the Surnames and Intermarriages of the Suburbanites of Shanghai. Acta Genetica Sinica,2002;29(8):666-673.

11. Cavalli-Sforza LL, Menozzi P, Piazza A. The History and Geography of Human Genes. Princeton: Princeton University Press,1996.

12. Bao MW: Analysis of Nanking Wu-Dialect during the Six Dynasty. In: Wu-Dialect Research Studio, Institute of Chinese Language and Literature. Collected Papers on Wu-Dialect. Shanghai: Shanghai Education Press, 1988; 15-17.

Fig.17   Photos of a Yue woman (left) and a Wu woman(right)