bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Han Chinese males with surnames related to the legendary *Huang* and *Yan* Emperors are enriched for the top two Neolithic super-grandfather Y chromosomes O3a2c1a and O3a1c, respectively

Pei He, Zhengmao Hu, Zuobin Zhu, Kun Xia, and Shi Huang*

State Key Laboratory of Medical Genetics

School of life sciences

Central South University

110 Xiangya Road

Changsha, Hunan, 410078, China

*Corresponding author: huangshi@sklmg.edu.cn

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

Most populations now use hereditary surnames, and most societies have patrilineal surnames. This naming system is believed to have started almost 5000 years ago in China. According to legends and ancient history books, there were Eight Great *Xings* of High Antiquity that were the ancestors of most Chinese surnames today and are thought to be descended from the two legendary prehistoric Emperors *Yan* and *Huang*. Recent work identified three Neolithic super-grandfathers represented by Y chromosome haplotypes, O3a1c, O3a2c1, and O3a2c1a, which makes it possible to test the tales of *Yan-Huang* and their descendant surnames. We performed two independent surveys of contemporary Han Chinese males (total number of subjects 2415) and divided the subjects into four groups based on the relationships of their surnames with the Eight Great *Xings*, Jiang (*Yan*), Ying (*Huang*), Ji(*Huang*), and Others (5 remaining *Xings* related to *Huang*). In both studies, we found that subjects with O3a1c were enriched with Jiang-related surnames and subjects with O3a2c1a were enriched with Ying-related surnames. Also, subjects with Jiang-related surnames were enriched with O3a1c and those with Ying-related surnames were enriched with O3a2c1a. Finally, subjects with O3a2c1 were slightly enriched for the Others-group, consistent with linking O3a2c1 to another legendary leader *Chi You* who lost to *Huang* and was largely ignored as an ancestor of Han on par with *Yan* and *Huang*. These results are remarkably consistent with historical writings on *Yan* and *Huang* and suggest that tales of *Yan-Huang* and their related-*Xings* and surnames may not be unrealistic.

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Introduction

The Han Chinese population uses hereditary surnames that are thought to be first established ~5000 years ago [1-3]. According to legends and ancient history books such as the Records of Grand History (*Shi-Ji*) by Sima Qian published at ~109 BC, modern Han Chinese are largely descended from *Yan-Di* (Yan Emperor) and *Huang-Di* (*Huang* Emperor) who lived ~5000 years ago. *Huang* defeated *Yan* and another legendary leader *Chi You* to lead the union of the three major tribes in China at the time. Prior to the Warring States period (fifth century BC), only the ruling families and the aristocratic elite had surnames. In ancient times two types of surnames existed, namely xing (Chinese: 姓; pinyin: xìng) or clan names, and shi (Chinese: 氏; pinyin: shì) or lineage names. Xing were surnames held by the noble clans. They generally are composed of a nü (女, "female") radical which has been taken by some as evidence they originated from matriarchal societies based on maternal lineages. There were “Eight Great *Xings* of High Antiquity” from ~4000 years ago that are believed to be ancestors of most of today’s ~23813 surnames of Chinese people [4]. Although these Eight Great *Xings* are thought to originate in matriarchal societies, it is still expected that certain males may be more dominant than others in such societies.

According to ancient history books like *Shi-Ji*, *Yan* belonged to one of the Eight Great *Xings*, Jiang, and had the surname Jiang that remains popular today. The other 7 Great *Xings* are all related to *Huang*. Of these, Ji is thought to be the original surname of *Huang* and has the most descendant surnames today. Ying is the surname given to Bo Yi (~2200 BC), widely known to be the 5th generation grandson of *Huang*, and has ~14 descendant surnames today. For the other five *Huang* related Great *Xings* (Ji2, Si, Wei, Yao, and Yun), each has less descendant surnames as well as less descendant populations today relative to Jiang, Ying, and

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Ji. Thus, *Huang's* descendants are expected to be enriched in surnames descended from either Ji or Ying.

Ying is special because one of its related contemporary surnames, Huang (meaning yellow), is also the same as the commonly known name for the Yellow Emperor or *Huang-Di* that literally means yellow soil and was originally inspired by the spirit of yellow soil in mainland China according to *Shi-Ji*. The character for soil was later changed to a different character for Emperor that has the same pronunciation and spelling, as calling a great leader soil seemed less respectful. It is expected by logic that *Huang'*s direct descendants should most likely use Huang as surnames rather than other lesser known names associated with him such as Ji, Gong Sun, and Xuan-Yuan that are very rare surnames today. Among the most famous Ying are Qin Shi Huang who united China at ~ 221BC and Laozi (~600 BC, surname Li) who invented the notion of *Dao* or *Tao* and authored one of the best-known Chinese philosophy books *Dao De Jing*. However, it remains to be established whether *Yan* and *Huang* as recorded by ancient history books actually existed, and if they did, whether the surnames claimed to be descended from them by ancient surname books such as "*Yuan He Surnames*" (812) are indeed so.

Recent progress on Y chromosome haplotypes makes it possible to test the tales of *Yan-Huang* and their related *Xings* and surnames. Consistent with Neolithic individuals matching the legendary status of *Yan*, *Huang* and *Chi You*, it has been found that there were three Neolithic super-grandfathers who claim ~40% of living Han Chinese males [5]. Their Y haplotypes originated ~5.4 Kya (thousand years ago) for O3a2c1a or Oα, ~6.5 Kya for O3a2c1-F444 or Oβ, and ~6.8 Kya for O3a1c-F11 or Oγ, and represent 16%, 11%, and 14% of present Han Chinese, respectively. Based on the estimated age and frequency, Oα could be a good candidate for *Huang* and Oγ for *Yan*. *Chi You*, who was defeated by *Huang,* could be a good candidate for the less popular Oβ, although not much is known regarding the descendant

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

surnames of *Chi You*. Therefore, we here tested whether contemporary Han males with surnames related to the Great *Xings* of Jiang and Ying are enriched with O3a1c and O3a2c1a, respectively.

Materials and Methods:

We collected and analyzed Y chromosome haplotype data of surnames representing 1564 unrelated Han males from the Website "One Surname a Week" (http://ranhaer.s47-56.myverydz.com/thread-29720-1-1.html) maintained by the researcher Dr. Shi Yan from Fudan University who was the first author on the three super-grandfathers paper [5]. The website put out on a nearly weekly basis genotyping results on Y haplotypes of the major surnames of Chinese males. The Fudan group has an active ongoing program to determine Y haplotypes for anyone from anywhere in China who with informed consent had sent their blood samples for genotyping and research analyses.

The grouping of contemporary surnames into the Eight Great *Xings* was based on popular surnames literatures [4]. We further divided the Eight Great *Xings* into four groups, Jiang, Ying, Ji, and Others (the remaining five *Xings*) because the ancient *Xings* in the Others-group each has relatively less descendant surnames. The fraction of each Y haplotype for each surname was calculated and the average fraction of all surnames in an ancient *Xing* group was used for group comparisons. Student's t test (one tailed) was performed to examine the hypothesis that O3a1c is more common in Jiang-group or that O3a2c1a is more common in Ying-group.

As an independent source of data, we collected peripheral blood samples from ~1200 unrelated Han males who were either individuals from Hunan area (~1000) or students (~200) from our campus in Hunan, China. The DNAs were extracted and genotyped by PCR-

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

sequencing for the three super-grandfather haplotypes. We determined for each individual Y sequence at positions Y:14105409 for O3a1c (A), rs202111911 for O3a2c1 (G), and Y:21399646 for O3a2c1a (A). The primers are: O3a1c, 5'-GGCTGAGATAATAGGGTCC and 5' –CCAGTTTTGGGCAAGAGTC; O3a2c1, 5'-GTCAAGTTCAGTGTGAGTTTTC, 5'-AATGTGTGGGTCCTTCAGTTTA; and O3a2c1a, 5'-TCCGACATAGGCAGAGGTATT, 5'-TGGAAGCTCAACATCCTTTTCAG. To replicate the results from the One Surname a Week data, we excluded surnames that were not present in that dataset or with too few samples. We also selected haplotype-defining SNPs equivalent to O3a1c-002611 rather than F11 or O3a2c1-M134 rather than F444 because data from the One Surname a Week used 002611 and M134 equivalent SNPs. F11 is the true haplotype for the super-grandfather Oγ but O3a1c should be a good approximate to represent it since it is the major branch under O3a1c. Similarly, F444 is the true haplotype for Oβ but O3a2c1 should approximate it. All samples were collected with informed consent and approved by the Biomedical Research Ethics Committee of the Central South University. Prior to DNA analysis, all samples were stripped of personal identifiers (if any existed). All procedures were in accordance with the ethical standards of the Responsible Committee on Human Experimentation (approved by the Biomedical Research Ethics Committee of the Central South University) and the Helsinki Declaration of 1975 (revised in 2000).

Results and Discussion:

We first made use of the Y chromosome haplotype data of surnames representing 1564 males as reported on the Website "One Surname a Week" maintained by researchers from Fudan University. These data did not type F11 and F444 but since they are major haplotypes underneath 002611 and M134, the available data on 002611 and M134 should approximate that

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of F11 and F444, respectively. Based on known phylogenetic relationships according to popular surnames literatures [4], we classified the surnames based on their relationships with the Eight Great *Xings* (Supplementary Table 1). To determine Y haplotype distribution among the Eight Great *Xings*, we divided contemporary surnames into 4 groups of Great *Xings*, Jiang, Ying, Ji, and Others (Supplementary Table 1). For each of the 4 groups, we obtained the average fraction of males per surname for each of the three super-grandfathers Y haplotypes (Figure 1). The Jiang-group has more O3a1c than the other groups (Jiang vs Ying, $P < 0.05$, Student's t test, one tailed). The Ying-group has more O3a2c1a than the other groups (Ying v Jiang, $P < 0.05$, Student's t test, one tailed). For the Jiang-group, O3a1c is more common than the other 2 haplotypes ($P < 0.01$, Student's t test, one tailed). For the Ying group, O3a2c1a is more common than the other two haplotypes, especially O3a2c1 (O3a2c1a vs O3a2c1, $P < 0.05$; Student's t test, one tailed). For both the Ji-group and the Others-group, O3a1c is slightly more common than the other two haplotypes ($P > 0.05$).

To verify the above result, we collected peripheral blood samples from subjects from the Hunan area in China and did PCR-sequencing genotyping on the 3 haplotypes. The results on 851 males again showed similar patterns of O3a1c enrichment in the Jiang-group, and O3a2c1a enrichment in the Ying-group (Figure 2).

The two independent data sets gave conflicting results regarding the Ji-group with one showing O3a1c enrichment while the other with O3a2c1a enrichment (Figure 1 and 2). Thus, if both the Ying and Ji group were descended from *Huang*, it is the Ying-group that has relatively less admixture with non-*Huang* lineages. The Others-group showed similar pattern as the Ji group with no consistent enrichment of any haplotype in the two datasets.

For males with the haplotype O3a2c1, both data sets showed slight enrichment of the Others-group of surnames (Figure 1 and 2). This group is less directly related to *Yan* and *Huang*

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

than Ying and Ji based on popularly known surname relationships. Therefore, it is likely that this group may be more related a legendary leader other than *Yan* and *Huang*. The only plausible candidate here would be *Chi You,* a leader of nearly equal status based on historical writings. Some Chinese history scholars have advocated for a Three-Ancestor or Emperor hypothesis for Han Chinese with *Chi You* included together with *Yan* and *Huang* (see for example, http://www.360doc.com/content/14/0107/20/5482926_343409144.shtml). The results here add weight to such a notion. Although the age of O3a2c1 was calculated to be ~1000 years earlier than *Huang*, this could be overestimated for two reasons. First, *Chi You* had lost the war and his people were known to flee to North-West and South-West mountains where they might have mated with local more primitive tribes, which would be expected to increase their Y allelic diversity due to mutations required for coevolution with mixed autosomes. Second, *Chi You* may be only the last of a series of tribal leaders who share the same paternal lineage.

The downplay of *Chi You* by ancient history books may simply reflect the fact that *Chi You* lost to *Huang-Di* and the common practice of some degree of self-serving in the winners' written account of history. The nearly absent account of a relationship between *Chi You* and contemporary Chinese surnames is a conspicuous anomaly inconsistent with *Chi You*'s status and may reflect the fact that people commonly want to associate their names with a victor. Thus, most *Chi You* descendants may have changed their surnames to be linked with *Huang* or were changed into names related to *Huang*. Consistently, even though males with O3a2c1 were enriched with surnames in the Others-group, the enrichment was very slight, unlike O3a1c with more prominent enrichment of Jiang-group or O3a2c1a with more prominent enrichment of Ying group (Figure 1 and 2). This pattern suggests that surnames originally linked to O3a2c1 had changed into surnames linked to *Yan* and *Huang* but the reverse did not happen as much. Also, a priori, of the 7 *Huang*-related Great *Xings*, other than Ying and Ji, the 5 *Xings* less directly

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

related to *Huang* within the Others-group here should have been more enriched with other lineages such as *Chi You*. And the DNA results here support such reasoning.

The legend of Han Chinese surnames deriving from *Yan* and *Huang* has few independent lines of evidence other than from a few ancient history books such as *Shi-Ji*. The association of the Jiang-group of surnames with O3a1c suggests *Yan* as the candidate ancestor of this haplotype. The book *Shi-Ji* says that Emperor *Yan* has surname Jiang and lived at a time earlier than *Huang*. The period of *Yan* is known to have dramatically expanded rice-related agriculture in the central south area of China and largely consistent with the age estimate for the haplotype O3a1c. The exact length of the *Yan* period is uncertain but would not be surprising if it is close to 500-1000 years or longer. It is also commonly believed that there were many different *Yan* Emperors who may share the same patrilineal lineage and the one who fought with and lost to *Huang* and later joined *Huang* to battle with *Chi You* may represent only the last of the *Yan* Emperors.

The association of the Ying surname with O3a2c1a suggests Emperor *Huang* as the candidate ancestor of this haplotype. Among the Eight Great *Xings*, Ying claims the second largest descendant population today. Five Ying-related surnames rank among the top 20 most popular surnames and together account for 13.9% Chinese based on a 2010 National Survey of China, second only to Ji-related surnames with 7 among the top 20 accounting for 23.8% Chinese and greater than Jiang-related surnames (2 among top 20 accounting for 2.8%) and the Others-group related surnames (4 among top 20 accounting for 9.4%) (Supplementary Table 3). There are many more Ji related surnames today than Ying-related ones and so it may be expected for Ji-related populations to be larger. That Ji-group of surnames has less representation of O3a2c1a than Ying-group indicates more admixture for Ji-group related populations, which are expected to increase surname diversity as well as population size.

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The Ying-related Huang surname, likely inherited most directly from Emperor *Huang*, is presently one of the most popular surnames (ranked 7th among all Chinese surnames). Most people with Huang surname trace their most recent common ancestry to Huang State near the southern border of Henan province in central China, which was an ancient major State founded at ~2148 BC by Bo Yi's descendants with their countrymen keeping the same name as their State's name after their country fell to the Chu State. However, the Huang surname seems not particularly highly enriched with O3a2c1a than the average (14% and 17% for Huang vs 11% and 17%, respectively, for the average of all non-Ying surnames in the two data sets here). Given the limited sample size in this study, we cannot determine with certainty the representative Y haplotype for each individual surname. In any event, the complex admixture history of each surname, which typically have >5 different ancestry stories, may make it unrealistic to identify a representative haplotype for most individual surnames. However, for a collection of surnames as a large group sharing a common ancestor, it may be possible to identify a common Y haplotype since the effect of random admixture or surname switching events may cancel each other out and have less impact on the overall pattern of a very large group. Our study here indeed indicates a meaningful assignment of representative Y haplotype for the Great *Xings* of Jiang and Ying.

The frequency of the three super-grandfather Y haplotypes is 0.17 for O3a1c, 0.12 for O3a2c1, and 0.14 for O3a2c1a in the 1564 samples data from "One surname a week" website, and 0.14, 0.098, and 0.16 respectively in the 851 samples data from this study (Supplementary Tables 1 and 2). These results are largely consistent with previous findings [5] and consistently rank O3a1c and O3a2c1a more popular than O3a2c1. Such a pattern is consistent with the Three-Ancestor or Emperor hypothesis where *Yan* and *Huang* as victors in their battle with *Chi You* are expected to leave more descendants than *Chi You*. To use 002611 and M134 rather than F11 and F444 to represent the super-grandfathers should overestimate the numbers a bit,

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

so the actual fractions of the super-grandfather *Yan* and *Chi You* should be a bit lower than those for O3a1c and O3a2c1, respectively. Since the *Huang*-related haplotype O3a2c1a is the youngest among the three, and yet has claimed similar if not more descendants than the *Yan*-related O3a1c, the oldest of the three, the pace of expansion for the *Huang* lineage appears to be the fastest among the three Y haplotypes. This is consistent with *Huang* being the ultimate victor among the three legendary leaders (*Huang* had defeated both *Yan* and *Chi You*). Together, these results here provide molecular evidence for the historical writings on *Yan* and *Huang* and the Eight Great *Xings* of High Antiquity.

Acknowledgements:

Supported by the National Natural Science Foundation of China grant 81171880 and the National Basic Research Program of China grant 2011CB51001 (S. H.).

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References:

1 Yuan, Y. D. Science and culture of surnames. *Chinese Nat. Geog.* **2**, 38-39 (2007).

2 Liu, Y., Chen, L., Yuan, Y. & Chen, J. A study of surnames in China through isonymy. *Am J Phys Anthropol* **148**, 341-350, doi:10.1002/ajpa.22055 (2012).

3 Jobling, M. A. In the name of the father: surnames and genetics. *Trends Genet* **17**, 353-357 (2001).

4 Yuan, Y. D. & Qiu, J. R. *Dictionary of Chinese Surnames*. (Jiangxi People's Press, 2010).

5 Yan, S. *et al.* Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers. *PLoS ONE* **9**, e105691, doi:10.1371/journal.pone.0105691 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure Legends:

Figure 1. Distribution of the three super-grandfather Y haplotypes among 1564 male subjects with surname and Y data from the website "One Surname a Week". The fraction of each haplotype in a surname was calculated and the average fractions per surname with standard error of the mean are shown in the plot.

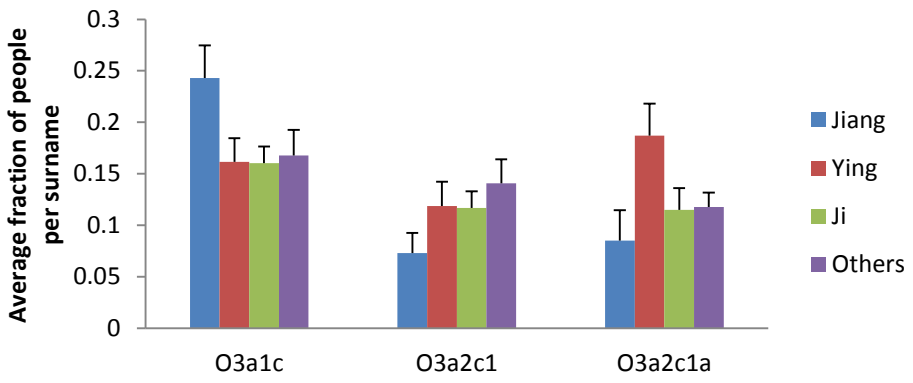
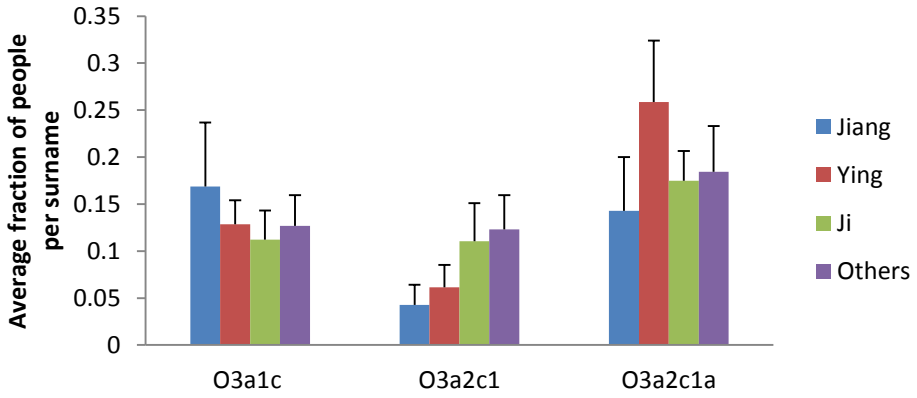


Figure 2. Distribution of the three super-grandfather Y haplotypes among 851 male subjects from this study. The fraction of each haplotype in a surname was calculated and the average fractions per surname with standard error of the mean are shown in the plot.



bioRxiv preprint doi: https://doi.org/10.1101/077222; this version posted September 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Supplementary Information:

Supplementary Table 1. Y chromosome haplotype distribution among Chinese surnames based on data from "One surname a week" website.

Supplementary Table 2. Distribution of the three Neolithic super-grandfather Y haplotypes among 851 male samples collected in this study.

Supplementary Table 3. Top 20 surnames from 2010 National Survey of China