

Development of an Okinawa panel for biogeographic inference of Okinawans

Nakanishi, Hiroaki; Pereira, Vania; Børsting, Claus; Tvedebrink, Torben; Takada, Aya; Saito, Kazuyuki

Published in:
Annals of human biology

DOI (link to publication from Publisher):
[10.1080/03014460.2023.2257594](https://doi.org/10.1080/03014460.2023.2257594)

Creative Commons License
CC BY 4.0

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Nakanishi, H., Pereira, V., Børsting, C., Tvedebrink, T., Takada, A., & Saito, K. (2023). Development of an Okinawa panel for biogeographic inference of Okinawans. *Annals of human biology*, 50(1), 436-441. <https://doi.org/10.1080/03014460.2023.2257594>

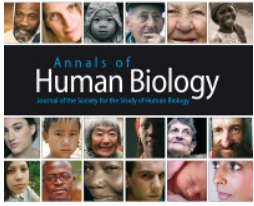
General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Development of an Okinawa panel for biogeographic inference of Okinawans

Hiroaki Nakanishi, Vania Pereira, Claus Børsting, Torben Tvedebrink, Aya Takada & Kazuyuki Saito

To cite this article: Hiroaki Nakanishi, Vania Pereira, Claus Børsting, Torben Tvedebrink, Aya Takada & Kazuyuki Saito (2023) Development of an Okinawa panel for biogeographic inference of Okinawans, *Annals of Human Biology*, 50:1, 436-441, DOI: [10.1080/03014460.2023.2257594](https://doi.org/10.1080/03014460.2023.2257594)

To link to this article: <https://doi.org/10.1080/03014460.2023.2257594>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



[View supplementary material](#)



Published online: 09 Oct 2023.



[Submit your article to this journal](#)



Article views: 64



[View related articles](#)



[View Crossmark data](#)

RESEARCH ARTICLE

OPEN ACCESS



Development of an Okinawa panel for biogeographic inference of Okinawans

Hiroaki Nakanishi^{a,b}, Vania Pereira^c , Claus Børsting^c , Torben Tvedebrink^d , Aya Takada^b and Kazuyuki Saito^{a,b}

^aDepartment of Forensic Medicine, Juntendo University School of Medicine, Tokyo, Japan; ^bDepartment of Forensic Medicine, Saitama Medical University, Saitama, Japan; ^cSection of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark; ^dDepartment of Mathematical Sciences, Aalborg University, Aalborg, Denmark

ABSTRACT

Background: The Precision ID Ancestry Panel with 165 SNP markers was unable to differentiate between mainland Japanese and Okinawa Japanese or to distinguish either of them from other East Asian populations.

Aim: An Okinawa panel was developed with the aim of further separating Okinawa Japanese individuals from mainland Japanese and other Asian groups. Seventy-five SNPs were selected using the most informative markers from the literature. Further, 22 SNPs were selected to separate Okinawa Japanese at minimum SNPs.

Subjects and methods: Samples were collected from 48 unrelated individuals from mainland Japan and 46 unrelated residents of the Okinawa prefecture. Data were evaluated by STRUCTURE, principal component, and GenoGeographer analyses.

Results: The 22 SNP set had similar levels of differentiation in STRUCTURE and PCA analyses as the 75 SNP set. GenoGeographer analysis showed that, out of the 46 Okinawa Japanese individuals, the 75 SNP and 22 SNP sets correctly assigned the Okinawan population as the most likely population of origin for 32 and 31 individuals, respectively.

Conclusion: Neither SNP set could completely differentiate between Okinawa Japanese and other Asian groups, however, these sets should be useful for crime investigation, when the sample, cost and time are limited.

ARTICLE HISTORY

Received 18 May 2023

Revised 6 July 2023

Accepted 31 August 2023

KEYWORDS

Mainland Japanese; Okinawan Japanese; single nucleotide polymorphism; ancestry inference; GenoGeographer

Introduction

Ancestry informative markers (AIMs) are useful for estimation of the biogeographic background of an individual and may be used to narrow down the number of possible suspects to a crime (Hollard et al. 2017; Sun et al. 2019). The development of next generation sequencing and commercial kits specifically designed for estimation of biogeographic ancestry (Jäger et al. 2017; Pereira et al. 2017) have made it easier to implement ancestry inference as a routine investigation in forensic genetic laboratories. Most commercial panels were developed for separation of main population groups, which means that intracontinental variation is usually not represented, and that these panels may not be suitable to distinguish among closely related populations. Therefore, specific sets of AIMs that can differentiate among populations within a specific region, such as North Africa and the Middle East populations (Pereira et al. 2019; Truelsen et al. 2021), European and South Asian populations (Phillips et al. 2013), East Asian countries (Shi et al. 2019), Korea (Jung et al. 2019) and Japan (Yuasa et al. 2015) have been reported.

The population structure of modern Japan is described by the dual structure model. The first inhabitants of the Japanese archipelago (Jomon people) came from Southeast Asia during the Upper Palaeolithic, approximately 30,000 years before current era (BCE). Historical evidence documents that this group admixed with a second group of migrants from Northeast Asia in the Yayoi age approximately 300 BCE (Yayoi people) (Hanihara 1991). Although the dual structure model has not been rejected completely, recent genetic analyses suggest that the Japanese population might have an even more complex structure (Jinam et al. 2015). The current Ainu population, which are indigenous groups living on the island of Hokkaido in the north most part of Japan, and the Ryukyuan population, which are indigenous groups living on the Okinawa Island in the south, are believed to retain more Jomon components based on morphological measurements (Jinam et al. 2015). Phylogenetic analyses of 641,314 SNPs showed that the mainland Japanese were located between an Ainu-Ryukyuan cluster and continental Asians, suggesting that they are an admixed population with components from these two populations (Jinam et al. 2012). Yamaguchi-Kabata

CONTACT Hiroaki Nakanishi hnakani@juntendo.ac.jp Department of Forensic Medicine, Juntendo University School of Medicine, 2-1-1, Hongo, Bunkyo-Ku, Tokyo, Japan

Supplemental data for this article can be accessed online at <https://doi.org/10.1080/03014460.2023.2257594>.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

et al. (2008) analysed 140,387 SNPs in 7,001 individuals from Japan and showed that the Ryukyu (another name for Okinawa) individuals formed a distinct cluster that was clearly separated from the mainland individuals. However, the large number of SNPs examined in the study is not realistic to be routinely applied in forensic genetic investigations. Nakanishi et al. (2018) investigated mainland Japanese and Okinawan Japanese using 165 SNPs from the Precision ID Ancestry Panel (Thermo Fisher Scientific, Waltham, MA) and performed population genetic analyses in the two populations. Their results showed that the panel could not differentiate between mainland Japanese, Okinawan Japanese, and other Asian populations. However, despite the lack of resolution for the targeted populations, several SNPs were identified that significantly differed amongst mainland Japanese and Okinawan Japanese.

Today, Okinawa is one of the most popular tourist sites in Japan and a lot of tourists from mainland Japan and nearby countries visit the islands. Moreover, many Americans live in Okinawa because there are several active US military bases on the main island. Therefore, in Okinawa prefecture, which is a multi-ethnic environment, the DNA profiling in criminal investigations should be important. Moreover, the sample, cost, and time are generally limited in a criminal investigation, so the ability to distinguish Okinawan Japanese using a small number of SNPs should be useful. In this work, we developed an SNP set to further differentiate between Okinawan and mainland Japanese. To reduce the number of SNPs needed for the investigations, we selected and evaluated the most suitable SNPs. Further, we created an "Okinawa-panel" with as few SNPs as possible, that may separate Okinawa Japanese individuals from mainland Japanese and other Asian groups.

Samples and methods

Samples and DNA extraction

Samples were collected from 48 unrelated individuals from mainland Japan living in 34 prefectures and 46 unrelated residents of the Okinawa prefecture. The parents of the individuals typed in this study were from the same prefecture. Genomic DNA was extracted from buccal swabs or FTA cards using the QIAamp DNA Mini Kit (Qiagen, Venlo, The Netherlands) following the recommendations of the manufacturer. Informed consent was obtained from the participants.

Selection and sequencing of SNPs

Supplementary Table S1 lists the 75 SNPs evaluated in this study, which were selected from the studies of Yamaguchi-Kabata et al. (2008) (25 SNPs), Nakanishi et al. (2018) (27 SNPs), Kamitsuji et al. (2015) (18 SNPs), and Yuasa et al. (2018) (5 SNPs). These SNPs were selected from SNPs showing (significant) differences in genotype frequencies between the mainland Japanese and Okinawa Japanese. Moreover, genetic distance for each SNP was also considered.

The 27 SNPs selected from the study of Nakanishi et al. (2018) were typed by the Precision ID Ancestry Panel based on methods in their study. The remaining 48 SNPs were typed by Sanger sequencing using the BigDye Terminator v. 1.1 Cycle Sequencing Kit (Thermo Fisher Scientific) and a 3730 Genetic Analyser (Thermo Fisher Scientific). The primer sequences are shown in Supplementary Table S2. PCR amplification of 48 SNPs was performed in 18 separate reactions containing 1–4 SNPs, as shown in Supplementary Table S2. Twenty microlitre reaction mixtures contained 10 μ L KOD One PCR Master Mix (Toyobo, Osaka, Japan), 1 μ L of 10 μ M oligonucleotide primers (final concentration = 0.5 μ M each), and 2 ng template DNA. Three-step PCR was performed using a SimpliAmp thermal cycler (Thermo Fisher Scientific) with the following program: 40 cycles of 98°C for 10 s, 60 or 62°C for 5 s, and 68°C for 1 s. Two-step PCR was performed using a SimpliAmp thermal cycle with the following program: 40 cycles of 98°C for 10 s and 68°C for 5 s.

Data analysis

Allele frequencies, Hardy–Weinberg Equilibrium (HWE), and Wright's Fixation index (F_{ST}) by locus-by-locus analysis of molecular variance (AMOVA) were calculated using Arlequin v3.5.2.2 (Excoffier and Lischer 2010). HWE analysis was carried out using 1,000,000 Markov Chain Monte Carlo (MCMC) steps and 1,000,000 dememorization steps. Pairwise linkage disequilibrium analyses and genetic differentiation among populations using Fisher's exact test were calculated for each SNP using Genepop v. 4.7.5 (Rousset 2008). Both tests were carried out using 10,000 MCMC steps, 1,000 batches, and 10,000 iterations per batch. Correction for multiple testing was performed according to Bonferroni (1936), by dividing the significance level of 0.05 by the number of comparisons.

STRUCTURE, PCA, and GenoGeographer interpopulation analyses were done using data from five Asian populations. In addition to the mainland Japanese (MNL) and Okinawa Japanese (OKI), reference data from five biogeographic regions (93 Chinese Dai in Xishuangbanna, China (CDX); 103 Han Chinese in Beijing, China (CHB); 105 Han Chinese South (CHS); 104 Japanese in Tokyo, Japan (JPT); and 99 Kinh in Ho Chi Minh City, Vietnam (KHV)) were collected from the 1,000 Genomes Project (<https://www.genome.gov/27528684/1000-genomes-project>). The distribution of genetic ancestry was investigated using STRUCTURE v. 2.3.4 (Pritchard et al. 2000; Falush et al. 2003) with 100,000 steps of burn-in followed by 100,000 repetitions for the MCMC. "Admixture" and "correlated allele frequencies" models were considered. The number of assumed populations (K) varied from $K=2$ to $K=6$ and three independent runs were performed for each value of K to check the consistency of the results. To determine the most likely value of K , Structure Harvester v. 0.6.94 was used (Earl and vonHoldt 2012), which implements the Evanno method (Evanno et al. 2005). The results of STRUCTURE were visualised using CLUMPP v. 1.1.222 (Jakobsson and Rosenberg 2007) and Distruct v. 1.1.23 software (Rosenberg 2003). PCA analysis was performed using an in-house script written in

Python based on the software `sklearn.decomposition.PCA v. 0.16.1` (Halko et al. 2011).

Each individual was further analysed using the GenoGeographer software (Tvedebrink et al. 2017; Mogensen et al. 2020), which was developed to assign the most likely population of origin and calculate the statistical weight of the evidence in the form of a likelihood ratio (LR). A z-score was computed for each individual, considering the seven Asian populations, including MNL and OKI, as reference. The analyses were performed using a leave-one-out approach, excluding the individual tested from the reference dataset each time the analysis was conducted. The population assignment was classified as “Rejected” if no reference population was assigned ($z\text{-score} > 1.64$, $p\text{-value} < 0.05$), “Ambiguous” if more than one metapopulation was assigned ($z\text{-scores} \leq 1.64$; $p \geq 0.05$) and the population likelihoods were not significantly different from each other ($z\text{-scores} \leq 1.64$; $p \geq 0.05$), or “Accepted” if only one metapopulation was assigned ($z\text{-score} \leq 1.64$; $p \geq 0.05$) or, if more than one metapopulation was assigned ($z\text{-score} \leq 1.64$; $p \geq 0.05$), one of the population likelihoods was significantly higher than the population likelihoods of all other assigned metapopulations ($p < 0.05$).

Results and discussion

The allele frequencies for the 75 loci in the mainland Japanese and the Okinawan Japanese are listed in [Supplementary Table S3](#). None of the loci were monomorphic in either population, and none of the 75 SNPs showed statistically significant deviation from HWE in either population ([Supplementary Table S4](#); $p\text{-value}$ after Bonferroni correction < 0.0007). No significant association between loci was found by pairwise LD testing (data not shown). [Supplementary Table S5](#) lists the numbers of genotypes in the mainland and Okinawan Japanese, and the $p\text{-values}$ by Fisher’s exact test for the 75 markers. Overall, 35 SNPs showed significant differences between the mainland and Okinawan Japanese. One of these SNPs was rs17822931-G/A (538G>A; Gly180Arg), located in the ABCC11 gene, that is responsible for determination of the earwax type (Yoshiura et al. 2006); AA homozygotes have dry ear wax, whereas AG heterozygotes or GG homozygotes have wet ear wax (Ohashi et al. 2011). In this study, the frequency

of the G allele (presented as the complementary allele, C on [Supplementary Table S3](#)) was significantly higher in the Okinawa Japanese than in the mainland Japanese ($p=0.006$). This is in agreement not only with a prior study (Yamaguchi-Kabata et al. 2008) but also with the report that this SNP has different allele frequencies among the Jomon and Yayoi people (Super Science High School Consortium 2009).

[Supplementary Table S6](#) shows the locus-by-locus AMOVA FST results and respective $p\text{-values}$ for the 75 markers. From these, a subset of 22 SNPs was selected based on an FST of $> 6.5\%$ and significant $p\text{-value}$ ($p < 0.05$) between the mainland Japanese and the Okinawan Japanese. STRUCTURE, Principal Component, and GenoGeographer analyses using the original 75 SNP set and the subset of 22 SNPs were performed. Besides the two Japanese populations, the analyses also included reference population data for other Asian populations from the 1000 genomes database (<https://www.genome.gov/27528684/1000-genomes-project>).

STRUCTURE analyses were run from $K=2$ to $K=6$ (data not shown). [Figure 1](#) shows the STRUCTURE results for the optimal $K=3$ detected by Structure Harvester for the 75 ([Figure 1A](#)) and 22 SNP ([Figure 1B](#)) sets. The global apportionment of ancestry for mainland and Okinawan Japanese is similar for both sets. The mainland Japanese present components shared by other Asian populations while the Okinawan Japanese seem to represent a separate cluster dominated by the blue component. The same was observed in the PCA plot ([Supplementary Figure 1](#)), where the Okinawan Japanese were more separated from the main cluster that included the other Asian populations (CDX, CHS, CHB, KHV, JPT, and mainland Japanese).

GenoGeographer analysis was done considering each of the seven Asian metapopulations. Each individual was investigated for their most likely population of origin, using a leave-one-out approach ([Table 1](#)). The 75 SNP set correctly assigned the Okinawan population as the most likely population of origin for 32 individuals (70%). Two individuals were rejected from the analysis, meaning that no relevant population of origin was represented in the reference population dataset. For the remaining 12 individuals, the population assignment was classified as ambiguous. This means that two or more reference populations were assigned as likely

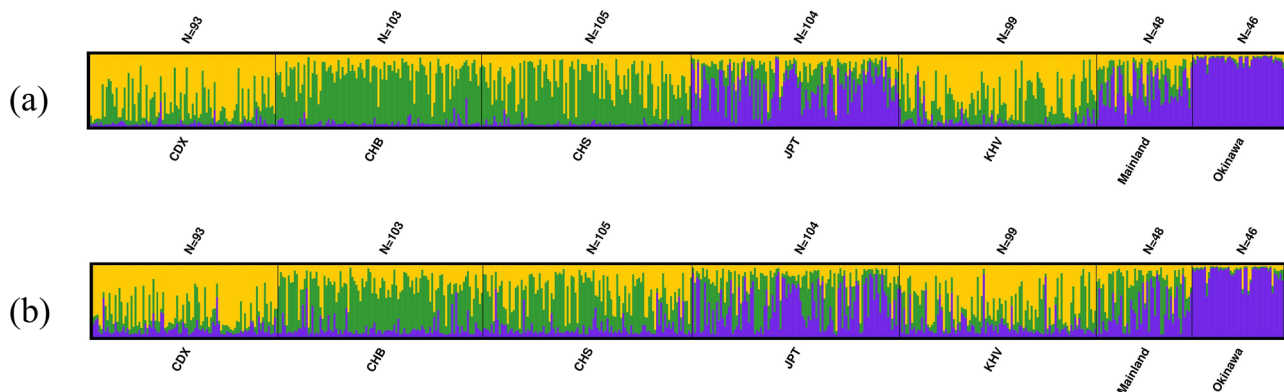


Figure 1. Membership coefficients by STRUCTURE analysis based on the (A) 75 and (B) 22 SNP sets. The most likely number of clusters (K) of both sets is 3.

Table 1. The results of the GenoGeographer analyses using the 75-SNP set (upper) and 22-SNP set (lower).

Origin	N	Rejected ^a	Ambiguous ^b	CDX	CHB	CHS	JPT	KHV	MNL	OKI
Chinese Dai in Xishuangbanna, China (CDX)	93	1 0	82 75	8 17	0 0	1 1	0 0	1 0	0 0	0 0
Han Chinese in Beijing, China (CHB)	103	1 0	91 89	1 2	7 11	2 1	0 0	0 0	1 0	0 0
Han Chinese South (CHS)	105	0 0	100 93	0 3	4 8	1 0	0 0	0 1	0 0	0 0
Japanese in Tokyo, Japan (JPT)	104	1 3	98 96	0 1	0 4	0 0	1 0	0 0	2 0	2 0
Kinh in Ho Chi Minh City, Vietnam (KHV)	99	1 1	91 80	2 8	0 6	1 0	0 0	4 4	0 0	0 0
Mainland Japanese (MNL)	48	1 1	43 43	0 1	0 1	0 0	0 1	0 0	4 1	0 0
Okinawa Japanese (OKI)	46	2 3	12 12	0 0	0 0	0 0	0 0	0 0	0 0	32 31

^aNo likely population among the reference populations. ^bAIM profile has at least two likely populations of origin and the likelihoods are not statistically significantly different.

Table 2. Median Log₁₀(LR)^a and range of LR^b for Okinawans.

Population B	Population A: Okinawans			
	Accepted (75 SNPs)	Ambiguous (75 SNPs)	Accepted (22 SNPs)	Ambiguous (22 SNPs)
Mainland Japanese	4.5 (2.2–8.2)	–	3.1 (0.9–6.5)	–1.2 (–1.0–1.3)
Han Chinese (Beijing)	14.1 (6.3–21.1)	6.1 (3.5–7.9)	6.4 (2.1–13.9)	2.4 (0.8–4.0)
Vietnam	11.3 (5.3–18.1)	4.3 (2.6–6.9)	5.8 (1.6–12.2)	2.1 (1.2–2.5)

^aLR = P(AIM profile|Population A)/P(AIM profile|Population B). LR was calculated if the two likelihoods were significantly different. ^bRange of LR^s are shown in parenthesis.

populations of origin, and that their likelihoods were not statistically significantly different. The 12 individuals with an “Ambiguous” classification were classified as originating from MNL, OKI, and/or JPT (Supplementary Figure 2). The 22 SNP set correctly assigned the Okinawan population as the most likely population of origin for 31 individuals (67%). Three individuals were rejected from the analysis. Seven of the remaining 12 individuals with an “Ambiguous” classification were classified as Japanese (any combination of MNL, OKI, and JPT, see Supplementary Figure 3).

The statistical weights of the population assignments were calculated in the form of likelihood ratios (Table 2) using the likelihoods reported by GenoGeographer. With the 75 SNP set, strong evidence (LR > 10,000) for Okinawan was obtained for Okinawan individuals with “Accepted” classification. Only when the alternative hypothesis was that the individual was mainland Japanese did 12 out of the 32 individuals obtain weight of the evidence that was less than 10,000. As expected, the LR^s were lower with the 22 SNP set. Nevertheless, LR > 10,000 for 11, 26, and 23 of the Okinawans with “Accepted” classification, when the alternative hypothesis was that the tested individual was MNL, CHB, or KHV, respectively. For Okinawans classified as “Ambiguous”, the LR^s were even lower, and only reached 10,000 or higher, when the alternative hypothesis was that the individual was CHB or KHV and when all 75 SNPs were used for the LR calculations.

Most of the mainland Japanese typed in this work were classified as ambiguous with both SNP sets, which suggested that this panel is not suitable for discrimination of mainland Japanese and other Asian populations. This was also reflected in the statistical weights of the population assignments

(Supplementary Table S7) that were lower than for the Okinawans and rarely reached 10,000. Furthermore, the likelihoods were not significantly different for many of the comparisons (25% and 41%, respectively, for the 75 and 22 SNP set) and LR^s were not calculated.

Although the weights of the population assignments of Okinawan individuals with the 22 SNP set was lower than with the 75 SNP set, separation of the Okinawan population from the other Asian groups could still be obtained with the smaller number of markers. The 22 SNP set may therefore be useful in cases where the sample material is limited. However, for most purposes, the 75 SNP set would be more useful for the identification of Okinawa Japanese. Nevertheless, more studies with other Asian populations and with more individuals from both mainland Japan and Okinawa are needed to confirm and validate the applicability of the two panels.

In conclusion, we created the Okinawa panel by selecting SNPs that differed significantly between the mainland and Okinawa Japanese. The Okinawa panel represents a novel SNP set towards the differentiation of Okinawan Japanese from mainland Japanese and other Asian groups (Yamaguchi-Kabata et al. 2008; Watanabe et al. 2021). Mizuno et al. (2021) reported that approximately 3,000 randomly selected SNPs were needed to distinguish Japanese subjects from other East Asians in a computer simulation. A panel with a small number of SNPs is unlikely to be able to completely differentiate between Okinawa Japanese and other Asian groups, however, the Okinawa panel should be useful for crime investigation, when the sample, cost, and time are limited.

Acknowledgements

We thank the Laboratory of Molecular and Biochemical Research, Biomedical Research Core Facilities, Juntendo University Graduate School of Medicine, for technical assistance. We are grateful to Mr. Manabu Uehara (Okinawa Prefectural Police HQ) and Mr. Shota Inokuchi (Juntendo University) for their cooperation with the collection of samples.

Disclosure statement

The authors declare that they have no conflicts of interest.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

ORCID

Vania Pereira  <http://orcid.org/0000-0002-8789-5286>

Claus Børsting  <http://orcid.org/0000-0003-0421-7429>

Torben Tvedebrink  <http://orcid.org/0000-0002-9292-8476>

Data availability statement

The datasets generated and/or analysed during the current study are available at Dryad, <https://doi.org/10.5061/dryad.5qfttdzbr>.

References

- Bonferroni CE. 1936. Teoria statistica delle classi e calcolo delle probabilità. Pubblicazioni del Regio Istituto Superiore di Scienze Economiche e Commerciali di Firenze. Vol. 8; Florence (Italy): Libreria Internazionale Seeber. p. 3–62.
- Earl DA, vonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour.* 4(2):359–361. doi: [10.1007/s12686-011-9548-7](https://doi.org/10.1007/s12686-011-9548-7).
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol.* 14(8):2611–2620. doi: [10.1111/j.1365-294X.2005.02553.x](https://doi.org/10.1111/j.1365-294X.2005.02553.x).
- Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 10(3):564–567. doi: [10.1111/j.1755-0998.2010.02847.x](https://doi.org/10.1111/j.1755-0998.2010.02847.x).
- Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics.* 164(4):1567–1587. doi: [10.1093/genetics/164.4.1567](https://doi.org/10.1093/genetics/164.4.1567).
- Halko N, Martinsson PG, Tropp JA. 2011. Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev.* 53(2):217–288. doi: [10.1137/090771806](https://doi.org/10.1137/090771806).
- Hanihara K. 1991. Dual structure model for population history of the Japanese. *Jpn Rev.* 2:1–33.
- Hollard C, Keyser C, Delabarde T, Gonzalez A, Vilela Lamego C, Zvenigorosky V, Ludes B. 2017. Case report: on the use of the HID-Ion AmpliSeq™ Ancestry Panel in a real forensic case. *Int J Legal Med.* 131(2):351–358. doi: [10.1007/s00414-016-1425-1](https://doi.org/10.1007/s00414-016-1425-1).
- Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, Walichiewicz P, Silva D, Pham N, Caves G, et al. 2017. Developmental validation of the MiSeq FGx Forensic Genomics System for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Sci Int Genet.* 28:52–70. doi: [10.1016/j.fsigen.2017.01.011](https://doi.org/10.1016/j.fsigen.2017.01.011).
- Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics.* 23(14):1801–1806. doi: [10.1093/bioinformatics/btm233](https://doi.org/10.1093/bioinformatics/btm233).
- Jinam T, Kanzawa-Kiriyama H, Saitou N. 2015. Human genetic diversity in the Japanese Archipelago: dual structure and beyond. *Genes Genet Syst.* 90(3):147–152. doi: [10.1266/ggs.90.147](https://doi.org/10.1266/ggs.90.147).
- Jinam T, Nishida N, Hirai M, Kawamura S, Oota H, Umetsu K, Kimura R, Ohashi J, Tajima A, Yamamoto T, et al. 2012. The history of human populations in the Japanese Archipelago inferred from genome-wide SNP data with a special reference to the Ainu and the Ryukyuan populations. *J Hum Genet.* 57:787–795.
- Jung JY, Kang P, Kim E, Chacon D, Beck D, McNevin D. 2019. Ancestry informative markers (AIMs) for Korean and other East Asian and South East Asian populations. *Int J Legal Med.* 133(6):1711–1719. doi: [10.1007/s00414-019-02129-7](https://doi.org/10.1007/s00414-019-02129-7).
- Kamitsuji S, Matsuda T, Nishimura K, Endo S, Wada C, Watanabe K, Hasegawa K, Hishigaki H, Masuda M, Kuwahara Y, et al. 2015. Japan PGx Data Science Consortium Database: SNPs and HLA genotype data from 2994 Japanese healthy individuals for pharmacogenomics studies. *J Hum Genet.* 60(6):319–326. doi: [10.1038/jhg.2015.23](https://doi.org/10.1038/jhg.2015.23).
- Mizuno F, Naka I, Ueda S, Ohashi J, Kurosaki K. 2021. The number of SNPs required for distinguishing Japanese from other East Asians. *Leg Med (Tokyo).* 49:101849. doi: [10.1016/j.legalmed.2021.101849](https://doi.org/10.1016/j.legalmed.2021.101849).
- Mogensen HS, Tvedebrink T, Børsting C, Pereira V, Morling N. 2020. Ancestry prediction efficiency of the software GenoGeographer using a z-score method and the ancestry informative markers in the Precision ID Ancestry Panel. *Forensic Sci Int Genet.* 44:102154. doi: [10.1016/j.fsigen.2019.102154](https://doi.org/10.1016/j.fsigen.2019.102154).
- Nakanishi H, Pereira V, Børsting C, Yamamoto T, Tvedebrink T, Hara M, Takada A, Saito K, Morling N. 2018. Analysis of mainland Japanese and Okinawan Japanese populations using the Precision ID Ancestry Panel. *Forensic Sci Int Genet.* 33:106–109. doi: [10.1016/j.fsigen.2017.12.004](https://doi.org/10.1016/j.fsigen.2017.12.004).
- Ohashi J, Naka I, Tsuchiya N. 2011. The impact of natural selection on an ABCC11 SNP determining earwax type. *Mol Biol Evol.* 28(1):849–857. doi: [10.1093/molbev/msq264](https://doi.org/10.1093/molbev/msq264).
- Pereira V, Freire-Aradas A, Ballard D, Børsting C, Diez V, Pruszkowska-Przybylska P, Ribeiro J, Achakzai NM, Aliferi A, Bulbul O, et al. 2019. Development and validation of the EUROFORGEN NAME (North African and Middle Eastern) ancestry panel. *Forensic Sci Int Genet.* 42:260–267. doi: [10.1016/j.fsigen.2019.06.010](https://doi.org/10.1016/j.fsigen.2019.06.010).
- Pereira V, Mogensen HS, Børsting C, Morling N. 2017. Evaluation of the Precision ID Ancestry Panel for crime case work: a SNP typing assay developed for typing of 165 ancestral informative markers. *Forensic Sci Int Genet.* 28:138–145. doi: [10.1016/j.fsigen.2017.02.013](https://doi.org/10.1016/j.fsigen.2017.02.013).
- Phillips C, Aradas AF, Kriegel AK, Fondevila M, Bulbul O, Santos C, Rech FS, Carceles MDP, Carracedo Á, Schneider PM, et al. 2013. Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries. *Forensic Sci Int Genet.* 7(3):359–366. doi: [10.1016/j.fsigen.2013.02.010](https://doi.org/10.1016/j.fsigen.2013.02.010).
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genet Soc Am.* 155(2):945–959. doi: [10.1093/genetics/155.2.945](https://doi.org/10.1093/genetics/155.2.945).
- Rosenberg NA. 2003. DISTRUCT: a program for the graphical display of population structure. *Mol Ecol Notes.* 4(1):137–138. doi: [10.1046/j.1471-8286.2003.00566.x](https://doi.org/10.1046/j.1471-8286.2003.00566.x).
- Rousset F. 2008. Genepop '007: a complete reimplementation of the genepop software for Windows and Linux. *Mol Ecol Resour.* 8(1):103–106. doi: [10.1111/j.1471-8286.2007.01931.x](https://doi.org/10.1111/j.1471-8286.2007.01931.x).
- Shi C, Liu Q, Zhao S, Chen H. 2019. Ancestry informative SNP panels for discriminating the major East Asian populations: Han Chinese, Japanese and Korean. *Ann Hum Genet.* 83(5):348–354. doi: [10.1111/ahg.12320](https://doi.org/10.1111/ahg.12320).
- Sun Q, Jiang L, Zhang G, Liu J, Zhao L, Zhao W, Li C. 2019. Twenty-seven continental ancestry-informative SNP analysis of bone remains to resolve a forensic case. *Forensic Sci Res.* 4(4):364–366. doi: [10.1080/20961790.2017.1306431](https://doi.org/10.1080/20961790.2017.1306431).
- Super Science High School Consortium. 2009. Japanese map of the earwax gene frequency: a nationwide collaborative study by Super Science High School Consortium. *J Hum Genet.* 54:499–503.
- Truelsen DM, Tvedebrink T, Mogensen HS, Farzad MS, Shan MA, Morling N, Pereira V, Børsting C. 2021. Assessment of the effectiveness of the EUROFORGEN NAME and Precision ID Ancestry panel markers for ancestry investigations of individuals from Europe, the Middle East, North Africa, North-East Africa, and South-Central Asia. *Sci Rep.* 11(1):18595. doi: [10.1038/s41598-021-97654-0](https://doi.org/10.1038/s41598-021-97654-0).
- Tvedebrink T, Eriksen PS, Mogensen HS, Morling N. 2017. GenoGeographer – a tool for genogeographic inference. *Forensic Sci Int Genet Suppl.* 6:e463–e465. doi: [10.1016/j.fsigs.2017.09.196](https://doi.org/10.1016/j.fsigs.2017.09.196).
- Watanabe Y, Isshiki M, Ohashi J. 2021. Prefecture-level population structure of the Japanese based on SNP genotypes of 11,069 individuals. *J Hum Genet.* 66(4):431–437. doi: [10.1038/s10038-020-00847-0](https://doi.org/10.1038/s10038-020-00847-0).

- Yamaguchi-Kabata Y, Nakazono K, Takahashi A, Saito S, Hosono N, Kubo M, Nakamura Y, Kamatani N. 2008. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. *Am J Hum Genet.* 83(4):445–456. doi: [10.1016/j.ajhg.2008.08.019](https://doi.org/10.1016/j.ajhg.2008.08.019).
- Yoshiura K, Kinoshita A, Ishida T, Ninokata A, Ishikawa T, Kaname T, Bannai M, Tokunaga K, Sonoda S, Komaki R, et al. 2006. A SNP in the ABCC11 gene is the determinant of human earwax type. *Nat Genet.* 38(3):324–330. doi: [10.1038/ng1733](https://doi.org/10.1038/ng1733).
- Yuasa I, Akane A, Yamamoto T, Matsusue A, Endoh M, Nakagawa M, Umetsu K, Ishikawa T, Iino M. 2018. Japaneseplex: a forensic SNP assay for identification of Japanese people using Japanese-specific alleles. *Leg Med (Tokyo).* 33:17–22. doi: [10.1016/j.legalmed.2018.04.008](https://doi.org/10.1016/j.legalmed.2018.04.008).
- Yuasa I, Umetsu K, Adachi N, Matsusue A, Nakayashiki N, Fujihara J, Akane A, Harihara S, Jin F, Ishikawa T. 2015. Investigation of Japanese-specific alleles most are of Jomon lineage. *Leg Med (Tokyo).* 17(1):52–55. doi: [10.1016/j.legalmed.2014.08.006](https://doi.org/10.1016/j.legalmed.2014.08.006).