Characterization of a novel private allele for East Asia and Oceania

J.L. Peterson[1], J.E. Ekins[1], K.L. Pennington[1], K.H. Ritchie[1], L.A.D. Hutchison[1,3], N.M. Myres[1], J.B. Ekins[1], J.E. Gomez[1], A.A. Nelson[1], M.E. Nelson[1], L. Layton[1], M.L. Lunt[1], T.A. Tolley[1], S.S. Masek[1], U.A. Perego[1],S.R. Woodward[1,2]. 1) Sorenson Molecular Genealogy Foundation (SMGF), Salt Lake City, UT, USA; 2) Dept. of Microbiology, Brigham Young University, Provo, UT, USA; 3) Computational Biology Research Group, CSAIL, Massachusetts Institute of Technology, MA, USA

Abstract

Genetic data as well as anthropological and linguistic evidence suggest that modern inhabitants of Oceania could have originated in a variety of world locations, including Asia. A large amount of the genetic data studied is based on either Y chromosomal or mitochondrial analysis (e.g. the "Polynesian Motif"), however, few of these studies have incorporated the X-chromosome (X-chr) as evidence of population origins. Herein we characterize a novel private allele on the X chr for the Oceanic region and East Asia. We have identified a 1-bp mutation in the STR locus DXS1041. 251 of 19,450 samples genotyped at that locus have this mutation, giving a frequency of .0129 in the SMGF database. Of these samples, 246 have X line genealogies deeply rooted in Asia or Oceania: 57.4% of these are Polynesian, 21.5% Asian, 14.6% mixed from both Asian and Oceanic populations, 2.8% New Zealander, 2.4% Southeast Asian and 1.6% Melanesian. Herein we propose that this allele could be used to identify membership in Asian/Oceanic populations.

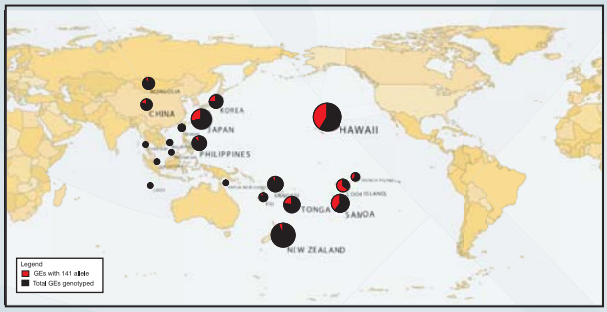

Figure 1. Geographic distribution of DXS1041 141 allele

Table 1

Region	Country	# NS*	# NS with 141 allele	Total Genetic Equivalents	Total 141 Genetic Equivalents	% with 141 allele
Asia	South Korea	18	5	19.75	5	25.3164557
	Taiwan	6	0	7.5	0.5	6.666666667
	Mongolia	37	4	37	4	10.81081081
	China	41	8	63.125	12.375	19.6039604
	Japan	138	37	149.375	39.625	26.52719665
Southeast Asia	Malaysia	2	0	2.25	0	0
	Indonesia	5	0	5	0	0
	Singapore	0	0	0.25	0	0
	Laos	1	0	1	0	0
	Thailand	3	0	3	0	0
	Phillipines	55	6	66.875	7.875	11.77570093
Melanesia	Fiji	6	1	6.25	1	16
	Papua New Guinea	1	0	1.25	0	0
	Vanuatu	40	3	40	3	7.5
Polynesia	Cook Islands	3	2	3	2	66.66666667
	French Polynesia	11	5	11.75	5	42.55319149
	Tonga	46	11	46.75	11	23.52941176
	Samoa	59	20	62.25	20.375	32.73092369
	Hawaii	245	106	302.375	124.5	41.17403886
Other	New Zealand	148	7	151.5	8	5.280528053
Totals	20	865	215	980.25	244.25	16.80677758

NS - Native Samples
* 4+ generations

Genetic Equivalents

Because of the complex inheritance pattern of the x-chromosome, units are given in genetic equivalents instead of the number of individuals. A genetic equivalent is defined as the percentage of an individual's X chromosomal genotype that is inherited from a specific ancestor. For example, a male inherits his one x-chr from his mother, who inherited half of her x-chr genotype (.5 genetic equivalents) from her mother and half from her father. His mother in turn inherited half of hers (.5*.5 = .25 g.e.s) from her mother and the other half (.5*.5 = .25 g.e.s) from her father [Figure 2]. The genetic equivalent values continue to divide as the X-line is traversed back. Since it would be inaccurate in this study to call, for instance, a .75 Hawaiian/.25 Chinese individual as a full Hawaiian, genetic equivalents were calculated for each individual [Table 1] and are used in Figure 1.

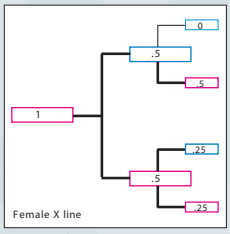

Figure 2. Genetic equivalents for the x chromosome of a female sample.

Conclusion

We have identified a potential private allele on the X chromosome that may be used to assign membership in Asian/Oceanic populations. The 141 allele in DXS1041 was observed in 24.9% of the Asian/Oceanic samples genotyped and in 1.29% of all samples in the SMGF database. Further research is needed to determine the probable origin and age of this mutation. Additional research could involve calculating MRCA values for the X-chr as well as investigating possible population substructure among these individuals.

Background

It is the general consensus of the anthropological community, as well as others, that modern day inhabitants of the Oceanic region originated in Asia and the surrounding area [Underhill 2001]. A variety of ideas, including the 'express-train to Polynesia' and the 'entangled-bank' hypotheses have been introduced to explain the colonization of the Pacific Islands. Current theories are based on data from mtDNA (e.g. the 'Polynesian Motif') and the Y chromosome [Kayser 2000]. We hope to expand the current methods to include data from the X chromosome by characterizing a potential private allele for the Asian/Oceanic region.

Description of Mutation

The potential private allele is a 1-bp mutation in the X-chr STR locus DXS1041, a 305-bp region in the lower section of Xq26. The individuals have the standard 16 CA repeats with an insertion of 1 thymine (T) 52 bp from the 3' end of the repeat, giving an allele value of 140 + 1 bp = 141 bp.

Populations

For our purposes, we have defined Oceania as the group of islands between the Coral Sea and the South Pacific Ocean, encompassing the eastern half of the island of New Guinea to the Hawaiian Islands. We have used the traditional definition of Polynesia as being those islands in the Hawaii-Tahiti-Tonga triangle. Islands from the Solomon Islands east to Fiji we have defined as Melanesia; countries south of mainland China extending to and encompassing the Philippines are defined as Southeast Asia.

Of the 19,450 total samples genotyped at DXS1041, 984 have documented X line ancestries deeply rooted in Asia and/or Oceania. 865 of these are what we will call "Native Samples;" meaning that their genealogies extend 4 generations or more into only one country [Table 1]. Admixture between Asian and Oceanic populations is observed in the other 119 samples [Table 2].

Table 2

Admixed Populations	# Samples
China/Japan*	4
China/Taiwan	2
China/New Zealand	1
China/Singapore	1
China/Malaysia	1
Phillipines/China	3
Phillipines/Japan	1
Phillipines/Hawaii	19
Phillipines/Caroline Islands	1
Phillipines/Hawaii/Japan	1
China/Phillipines/Hawaii	1
Samoa/Fiji/Papua New Guinea	1
Hawaii/Samoa	4
Hawaii/China	39
Hawaii/Japan	23
French Polynesia/Samoa	1
Hawaii/Korea	4
Hawaii/Taiwan	1
Hawaii/French Polynesia	1
Samoa/China	1
Hawaii/China/Korea	1
Samoa/New Zealand	2
New Zealand/Tonga	1
Hawaii/New Zealand	3
China/Japan/Hawaii	1
Tonga/Hawaii/Phillipines	1
Total	119

* Countries with the greatest genetic equivalent are named first

251 total samples were observed in the SMGF X-chromosome database as having the 141 allele. 246 are of Asian and/or Oceanic descent, with 5 samples of US/European origins having no evidence of any Asian or Oceanic influence.

These give a frequencies of .0129 for the potential private allele in the SMGF X chromosome database and .25 in the regions of interest. Of the 251total samples with this allele, 57.4% are native Polynesian, 21.5% are native Asian, 2.8% native New Zealander, 2.4% native Southeast Asian, 1.6% native Melanesian, and 1.9% are of US/European origins.

Sequence Analysis

One 4-generation Japanese male and one 4-generation Samoan male, both with the 141 allele, were sequenced. One Hawaiian male, with the CEPH allele value of 140, was also sequenced as a control. Sequences were analyzed on the program SEQUENCHER. The control sample was the same as the reference sequence, giving both allele values of 140. The Japanese and Samoan samples also had the same sequence as the reference, however, an insertion of one thymine was detected 52 base pairs downstream from the end of the CA repeat segment in both samples. This 1-bp insertion, along with the 16 CA repeats, gives the samples an allele value of 141. Sequence results are given in Figure 3.

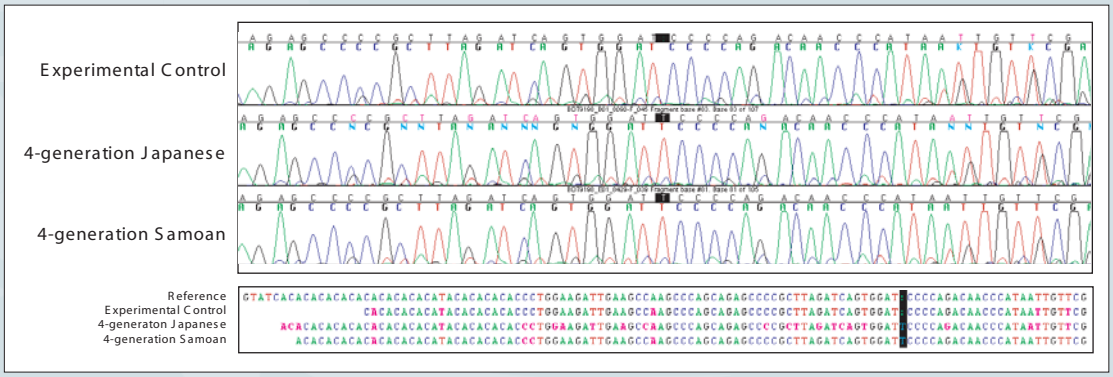

Figure 3. STR sequences for 3 samples

Acknowledgements

Special thanks to all those who contributed their DNA and pedigrees to the Molecular Genealogy Research Project (MGRP) and to those who assisted with sample collections. Thanks to Helaman Escobar with the University of Utah DNA Sequencing Core for STR sequencing and to Edgar Gomez for assistance with formatting. This study was supported by the Sorenson Molecular Genealogy Foundation, a non-profit organization, and its contributors.

References

Falush, D. et al. (2003). Genetics 164:1567-1587.
Harris E.E. et al. (1999) Proc. Natl. Acad. Sci. USA 96: 3320-3324.
Kayser, M. et al. (2000). Curr Biol 10: 1237-1246.
Laan M. et al. (2005). Eur J Hum Genet 13: 452-462.
Schaffner, S.F. (2004). Nat Rev Genet 5: 43-51.
Underhill, P.A. et al. (2001). Hum Mut 17:271-280.
http://www.ncbi.nlm.nih.gov/Genbank

SORENSON MOLECULAR GENEALOGY FOUNDATION
www. smgf.org