



# 太平洋隔离群体萨摩亚人的遗传结构分析

奚慧峰, 代卡·兰江

辛辛那提大学环境健康系, 美国 俄亥俄州 辛辛那提 46267

**摘要:** 萨摩亚人群是太平洋上一个有 3000 年历史的隔离岛屿群体。对萨摩亚人基因组的研究表明它拥有比较单纯的遗传背景和高连锁不平衡水平, 并且支持波利尼西亚人群起源的“快车模型”。随着东西萨摩亚的分治, 东西萨摩亚的生活饮食方式已经有了巨大的差异, 很多疾病的发病率也有很大的差别, 但是两者在遗传上仍然是一个统一的群体。所有这些都使得萨摩亚人群成为一个复杂疾病研究的良好对象。

**关键词:** 西萨摩亚; 美属萨摩亚; 波利尼西亚; 隔离群体; 复杂疾病

## Genetic Structure Analysis of Samoan: an Isolated Population in the Pacific

XI Huifeng, Ranjan DEKA

Department of Environment Health, University of Cincinnati, Cincinnati, OH45267 USA

**ABSTRACT:** Samoan is an isolated population inhabiting in South Pacific for more than 3000 years. A consistent genetic background and a high linkage disequilibrium level are found through the analysis of Samoan genome. Genetic evidences are more supportive for the “Express Train” model of Polynesian origin. There are large differences in living style and incidence rates of various diseases between West Samoan and American Samoan. However, these two populations are genetically homogeneous. All these facts make Samoan to be a model population for complex disease studies.

**Key words:** West Samoan; American Samoan; Polynesian; Isolated population; Complex disease

萨摩亚群岛(Samoan Islands)位于南太平洋, 地理上位于波利尼西亚群岛的中心地带, 有“波利尼西亚的心脏”之称。萨摩亚群岛是火山岛, 其人类的定居史可以追溯到公元前 1000 年[1]。从 1899 年开始, 根据协议德国和美国分割萨摩亚群岛, 美国占领了群岛的东半部, 德国占领了西半部。1962 年 1 月 1 日, 西萨摩亚成为以阿皮亚(Apia)为首都的萨摩亚独立国(Independent State of Samoa)。而东萨摩亚仍然是美属萨摩亚(American Samoa), 是美国领土的一部分。东西萨摩亚都以萨摩亚人(属于波利尼西亚人种)为主要人群, 辅以少量欧洲人群和杂合人群。西萨摩亚人口约 17 万(2004 年统计), 社会结构仍然是部落制和农耕文明为主; 东萨摩亚人口约 6.5 万(2005 年统计), 工业化相当普遍, 生活方式已经接近美国本土。

## 人类学历史

大约 3000 年前, 波利尼西亚文明就已

经建立并且以 Tonga 和萨摩亚为中心繁荣起来。现在关于波利尼西亚人群的定居模式的假说, 主要有“快车模型”(Express Train)和“混合库模型”(Entangled Bank)。其中快车模型的主要设想是大约 4000-5000 年前, 从东南亚的农业文明的核心地带出发, 人类发生了一次快速的向东迁移, 遍及大量的太平洋岛屿, 并于 3000 年前左右到达波利尼西亚群岛地区(见图 1)。一些线粒体 DNA 的数据支持这种学说[2,3]。而“混合库模型”假说由 Terrel[4]提出, 核心观点是萨摩亚人主要来自于附近的美拉尼西亚, 经过长期反复的交流演化才形成今天的波利尼西亚人群。现今的分子遗传学证据, 多数支持快车模型。比如, 宿兵等在 2000 年报道[5], 通过 Y 染色体的单倍型研究, 在研究的全部 17 种单倍型(H1~H17)中, 萨摩亚群体只拥有其中的 4 种 H1、H5、H6、H14, 分别对应原始的单倍型和突变 M130、M9、M122、M45。研究并没有发现波利尼西亚人群中拥有美拉尼西亚人群特异性的单倍型。其他的研究也

收稿日期: 2007 年 6 月 18 日 修回日期: 2007 年 6 月 24 日 联系人: 奚慧峰 xihufeng@gmail.com

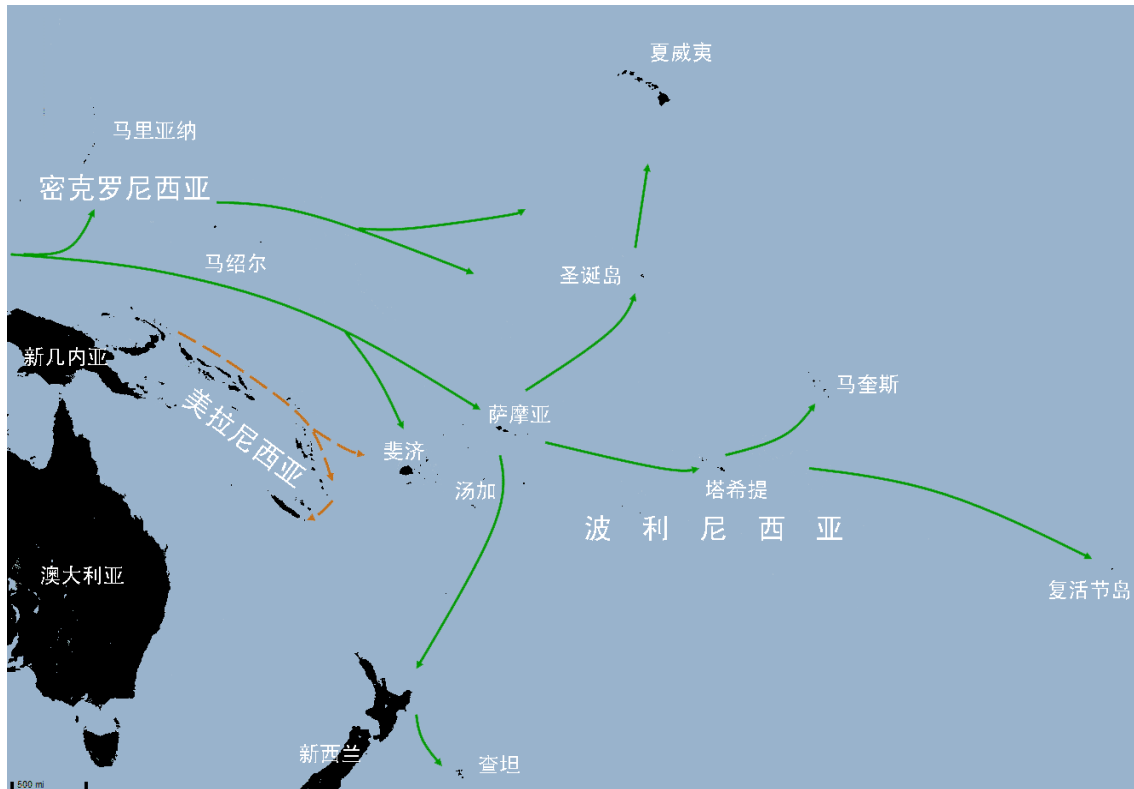


图 1 根据“快车模型”绘制的美拉尼西亚和波利尼西亚人群迁徙示意图

Fig.1 Migration routes of Melanesian and Polynesian according to “Express Train” hypothesis

证实了与其他标准人群相比，萨摩亚群体存在低的杂合度和高的背景连锁不平衡值[6,7]。这些证据可以推测出萨摩亚群体的发展经过了一个明显的瓶颈效应，而不是一个反复混合的过程。对于萨摩亚祖先群体大小的估计一直在进行当中，虽然要准确地估计数目非常困难，但是有理由认为当初最先在萨摩亚定居的人数非常少。而事实上在 18 世纪西方人发现萨摩亚群岛之前，由于优越的自然环境，人口的增长一直稳定而迅速。西方人到达萨摩亚之后，则由于因为引入了新的疾病的缘故，人口一度减少，直到 20 世纪才恢复增长[8]。

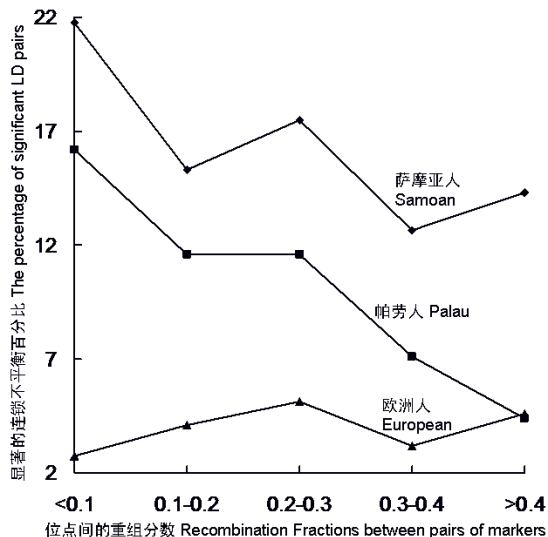
## 群体遗传结构

虽然萨摩亚分为西萨摩亚和美属萨摩亚两个独立的行政体，但是相互之间的通婚和交流事实上从来也没有停止过。从遗传结构上来看，两者仍然是一个统一的群体。针对西萨摩亚和美属萨摩亚混合样本，无论是以 STR(Short Tandem Repeat 短片段重复)为材料的较大规模针对性群体结构检测，还是附

带性质的 Genome Control 的方式来校正群体结构的相关分析研究项目，都没有发现明显的群体亚结构。主要的研究列举如下：1) Deka 等 1994 年报道了用 9 个 HVR (hypervariable region 高变区)位点和 4 个血清蛋白多态位点进行的东萨摩亚人的比较研究表明，绝大部分等位基因都是东萨摩亚人群共享的，频率上也没有明显的差异[6]；2) Tsai 等 2000 年报道了用了 25 个非连锁的 STR 位点，用群体遗传结构检测软件 STRUCTURE[9]对 400 多个分别来自东萨摩亚的个体进行检测，没有发现任何群体亚结构[7]。3) 利用全基因组范围内 300 多个有一定连锁关系的 STR 位点，用新版的可以处理连锁关系的 STRUCTURE 软件[10]，仍然没有发现任何群体亚结构(Deka 等未发表)；4) 选用 100 个非连锁的 SNP 位点，进行 Genome Control 为基础的相关分析的因子校正，计算出的校正因子是 0.96，非常接近完全没有群体亚结构的 1.00 [11]。这些结果反复证明了东萨摩亚在遗传结构上的一致性，从而支持东萨摩亚作为一个整体进行遗传研究。

## 隔离群体

由于岛屿地理原因，定居于萨摩亚群岛的人群一直是一个很好的隔离群体的例子。隔离群体的证据除了当地记录和历史记载，基因组遗传标记的研究也提供了有力的支持。萨摩亚群体与其他一般群体相比，可以看到明显的更低的杂合度，更缓慢的连锁不平衡衰退和更高的连锁不平衡水平[7]（见图2）。而采用 Affymetrix 250k 的基于 SNP 的连锁不平衡研究中，虽然在 30k 左右范围内，连锁不平衡水平和其他 Hapmap 群体差别很小，但是对于更长的距离范围内的连锁不平衡有显著的提高，基因组的杂合度水平也比较低(Deka 等未发表)。



**图2 萨摩亚人群与帕劳和欧洲白种人群相比的连锁不平衡水平** 改编自[7]。横坐标是位点间的距离(用重组分数来表示),纵坐标是对应距离上位点间的显著的连锁不平衡所占的百分比。在所有距离上,相比帕劳和欧洲白种人,萨摩亚群体均显示了最高的连锁不平衡水平。

Fig.2 The comparison of Linkage Disequilibrium among Samoan, Palau and European (Adapted from [7]). The X-axis is the recombination fractions between pair of markers. The Y-axis is the corresponding percentage of significant LD pairs. Compared to Palau and European, the highest level of LD in Samoan is observed on all distances.

随着 HapMap 计划的完成, Tagging SNP 的思路便利了疾病基因的寻找。事实上,由于萨摩亚群体与 HapMap 中的中国人群在遗传距离上并不远,且共享了大量的连锁不平衡,直接采用基于 HapMap 中国人群的 Tagging SNP,在萨摩亚人群样本中能覆盖到大约 85% 的未分型位点,这一比例是令人满意的[12]。

## 复杂疾病研究

隔离群体与一般群体相比,进行遗传疾病相关分析的时候,有很多的优势:更单纯的遗传背景,更高的连锁不平衡水平,更一致的生活环境等等。所以萨摩亚群体是一个研究复杂疾病的良好材料。更有意思的是东西两个人群的生活方式差异的发生。原来的萨摩亚人群生活方式比较原始,随着东西萨摩亚的分治,二战之后,东萨摩亚人的生活方式逐渐现代化,饮食结构大大改变而接近西方习惯,而西萨摩亚人的饮食和生活仍然保持传统。结果导致东萨摩亚人的一些代谢疾病,比如肥胖,发病率远远高于一般群体,也大大高于西萨摩亚人[13]。东萨摩亚人女性有 84%,男性有 78% 的肥胖率,而西萨摩亚这 2 个比例分别是 59% 和 29%。同样的,对于 II 型糖尿病,东萨摩亚女性发病率高达 25%,男性也高达 15%,而西萨摩亚的发病比例仅仅是 3% 和 5%[11]。这些特点支持 Neel 于 1962 年提出的“节俭”基因的假说[14],同时证明了萨摩亚人群有很高比例的易感基因。所谓的“节俭”基因,就是指某些基因形式有利于在不利的生存条件下提高生物的生存能力(比如更有效的储存和利用糖类和脂肪),但是当条件优越的时候反而成为某些代谢疾病的致病基因。东西萨摩亚人群巨大的患病率差异和相似的遗传背景结合起来,既有利于准确定位这些基因,也大大便利了相关的研究,比如疾病和环境的关系。McGarvey 等 2002 年报道了 LEP(Leptin)基因和萨摩亚人群 BMI(body mass index 体重指数)的相关性[15];同时对神经肽代谢相关基因和肥胖的关系,以及糖尿病相关基因的寻找,也在萨摩亚人群中积极开展[11]。

## 总结和展望

萨摩亚的独特地理环境和人类学历史，使萨摩亚人群成为一个单一起源的典型隔离群体。而与其他隔离人群相比，萨摩亚人群拥有更加优越的研究特点：高的代谢类疾病的发病率，和东西萨摩亚分治带来的不同的生活环境和饮食习惯的影响。在接下来的几年中，可以预见以萨摩亚人群为对象的肥胖，糖尿病，高血压研究都会取得相当的进展。也会对隔离群体在复杂疾病研究中的意义提供更多的佐证。同时，关于萨摩亚人群起源和与周围人群的关系的研究也将更加细致而深入。

### 参考文献

- Kirch PV (2000) On the road of the Winds, an archaeological history of the pacific islands before european contact. University of California Press. Berkeley CA.
- Sykes B, Leiboff A, Low-Beer J, Tetzner S, Richards M (1995) The origins of the Polynesians: an interpretation from mitochondrial lineage analysis. *Am J Hum Genet* 57:1463-75.
- Redd AJ, Takezaki N, Sherry ST, McGarvey ST, Sofro AS, Stoneking M (1995). Evolutionary history of the COII/tRNA intergenic 9 base pair deletion in human mitochondrial DNAs from the Pacific. *Mol Biol Evol* 12:604-615.
- Terrel J (1988) History as a family tree, history as an entangled bank: constructing images and interpretations of prehistory in the South Pacific. *Antiquity* 62:642-657.
- Su B, Jin L, Underhill P, Saha N, McGarvey ST, Shriver MD, Chu J, Oefner P, Chakraborty R, Deka R (2000) Polynesian origins: Insights from the Y chromosome. *Proc Natl Acad Sci U S A* 92:8225-8228.
- Deka R, McGarvey ST, Ferrell RE, Kamboh MI, Yu LM, Aston CE, Jin L, Chakraborty R (1994) Genetic characterization of American and Western Samoans. *Hum Biol* 66:805-822.
- Tsai H, Sun G, Smelser D, Viali S, Tufa J, Jin L, Weeks DE, McGarvey ST, Deka R (2004) Distribution of genome-wide linkage disequilibrium based on microsatellite loci in the Samoan population. *Hum Genomics* 1: 327-334.
- McArtlur N (1967). *Island Populations of the Pacific*. Canberra: Australian National University Press.
- Pritchard JK, Stephens M, Donnelly PJ. (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945-959.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics* 164: 1567-1587.
- Smelser D (2006) A Comparison of Obesity Candidate Genes in the Anabolic Neuropeptide Pathway in the Samoan and American Samoan Populations. University of Cincinnati PhD Thesis. 1-144.
- Deka R, Huang W, He Y, Wang H, Wang Y, Wang Y, Li H, Weeks DE, McGarvey ST, Jin L (2005) Portability of the HapMap: Linkage Disequilibrium sharing between the continental populations and the Samoans of Polynesia. ASHG Annual Meeting, Salt Lake City, UT, October 23-29, 2005.
- Galanis DJ, Sobal J, McGarvey ST, Pelletier DL, Bausserman L (1995) Ten-year changes in the obesity, abdominal adiposity, and serum lipoprotein cholesterol measures of Western Samoan men. *J Clin Epidemiol* 48:1485-1493.
- Neel JV (1962). Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am J Hum Genet* 14: 353-362.
- McGarvey ST, Forrest W, Weeks DE, Sun G, Smelser D, Tufa J, Viali S, Deka R (2002) Leptin locus (LEP) alleles and BMI in Samoans. *Int J Obes Relat Metab Disord* 26:783-788.