SMRT Sequencing of Whole Mitochondrial Genomes and Its Utility in Association Studies of Metabolic Disease

Adam English[1], Arnaud Besse[1], Min Wang[1], Yi Han[1], Vanessa Vee[1], Yan Gao[2], Jeff Reid[1], Donna Muzny[1], Richard A. Gibbs[1], Ellen E. Paxinos[2], Penelope E. Bonnen[1]

1. Baylor College of Medicine, Department of Molecular and Human Genetics, Human Genome Sequencing Center 2. Pacific Biosciences

ABSTRACT:

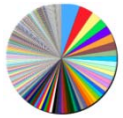
In this study we demonstrate the utility of Single-Molecule Real Time SMRT™ sequencing to detect variants and to recapitulate whole mitochondrial genomes in an association study of Metabolic syndrome using samples from a well-studied cohort from Micronesia. The Micronesian island of Kosrae is a rare genetic isolate that offers significant advantages for genetic studies of human disease. Kosrae suffers from one of the highest rates of MetS (41%), obesity (52%), and diabetes (17%) globally and has a homogeneous environment making this an excellent population in which to study these significant health problems. We are conducting family-based association analyses aimed at identifying specific mitochondrial variants that contribute to obesity and other co-morbid conditions. We sequenced whole mitochondrial genomes from 10 Kosraen individuals who represent greater than 25 % of the mitochondrial genetic diversity for the entire Kosraen population. Using Pacific Biosciences® C2 chemistry, SMRTbell™ libraries were constructed from pooled, full-length, unsheared 5 kb PCR amplicons, tiling the entire 16.6 kb mtDNA genome. Average read lengths for each sample were between 2500-3000bp, with 5% of reads between 6,000-8,000 bases, depending on movie lengths. The data generated in this study serve as proof of principle that SMRT sequencing data can be utilized for identification of high-quality variants and complete mitochondrial genome sequences. These data will be leveraged to identify causative variants for Metabolic syndrome and associated disorders.

INTRODUCTION:

Kosrae has 7,700 current residents on the island, 3,200 of whom enrolled in this study. The vast majority of individuals in this cohort comprise one pedigree (N=3031) and virtually all individuals mitochondrial genomes present in this population are members of mitochondrial haplogroups M and B. Higher resolution genetic analysis is required in order to distinguish the mitochondrial genetic contribution to disease. Whole mitochondrial genome sequencing is currently being pursued.

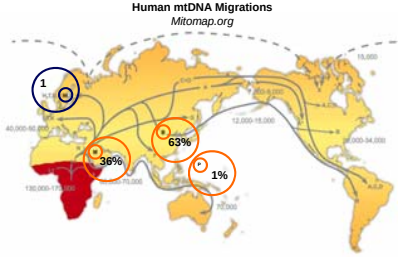
Maternal Lineages in Kosraen Pedigree

- 178 (N≥3) maternal lineages on island; 48 (N≥25)
- Top 10 largest lineages account for 28% of population
- 99% haplogroup M or B



Human mtDNA Migrations

Mitomap.org



Mitochondrial haplogroup genotyping of 3193 Kosraens

RESULTS:

Association study for mitochondrial variants contributing to metabolic disease on Kosrae

- Compare maternal lineage to non-maternal lineage in one pedigree (Wilson *et al* 2004)
- 20 mitochondrial lineages are present in pedigrees that possess ≥ 30 maternal and non-maternal members
- Association study for qualitative and quantitative traits: Metabolic Syndrome, diabetes, BMI, waist, triglycerides, total cholesterol, blood pressure

Table 1. Association study shows significant difference in BMI and waist between the mt and non-mt members of five Kosraen pedigrees.

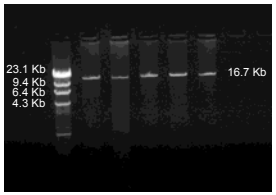
			BMI			Waist		
Founder	MT N	NONMT N	MT Mean	NONMT Mean	P-value	MT Mean	NONMT Mean	P-value
1	168	191	29	31	1.6E-12	89	93	1.7E-07
2	89	206	28	30	1.9E-12	89	92	5.9E-05
3	34	106	28	30	5.1E-10	86	92	3.6E-07
4	100	280	30	31	3.2E-08	91	94	2.7E-06
5	66	392	29	30	1.8E-06	88	91	1.9E-07
6	67	57	31	29	1.5E-05	93	90	5.5E-02
7	42	85	30	32	7.6E-05	91	96	2.4E-04
8	29	26	31	30	4.0E-04	93	89	4.1E-04
9	27	19	32	34	4.8E-03	97	104	1.6E-06
10	24	81	32	31	5.0E-03	95	94	3.3E-01

Sequencing Approach

Control individual GM12004 was sequenced via two approaches to determine the best sequencing method for assessing mitochondrial genomic variants. The entire mt genome was amplified with one 16.7 Kb PCR amplicon. This amplicon was sequenced by two different approaches:

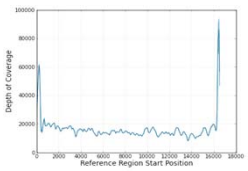
1.500 bp insert library sequenced using early release C2 chemistry (ERC2) at BCM-HGSC.

2.Direct sequencing of full length amplicon using C2 chemistry at PacBio.

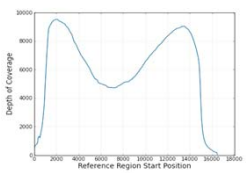


Coverage

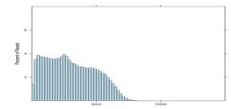
ChrM Coverage (500 bp insert)



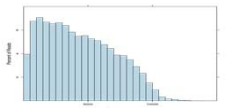
ChrM Coverage (16.7 Kb insert)



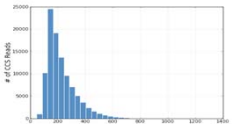
BCM HGSC 500 bp insert library | **PacBio 16.7 Kb insert library**



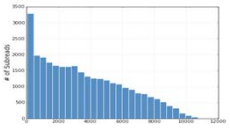
Read Length



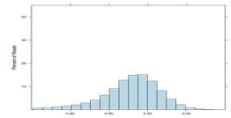
Read Length



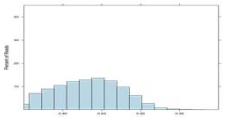
Mapped SubRead Length



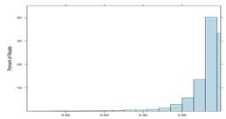
Mapped SubRead Length



Accuracy



Accuracy

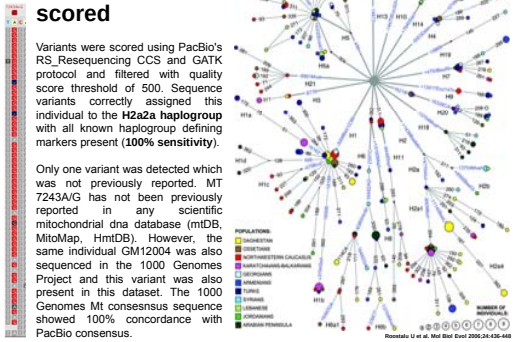


CCS Accuracy

Mitochondrial variants accurately and sensitively scored

Variants were scored using PacBio's RS_Resequencing CCS and GATK protocol and filtered with quality score threshold of 500. Sequence variants correctly assigned this individual to the **H2a2a haplogroup** with all known haplogroup defining markers present (**100% sensitivity**).

Only one variant was detected which was not previously reported. MT 7243A/G has not been previously reported in any scientific mitochondrial dna database (mtDB, MitoMap, HmtDB). However, the same individual GM12004 was also sequenced in the 1000 Genomes Project and this variant was also present in this dataset. The 1000 Genomes Mt consesnsus sequence showed 100% concordance with PacBio consensus.



Roostalu U et al. Mol Biol Evol 2006;24:436-448

Conclusions

SMART sequencing using 500 bp inserts accurately and sensitively identifies variants in the mt genome allowing for high resolution of genetic variants.

Acknowledgements

The authors acknowledge Lawrence Lee, Jason Chin, David Villagos for their support of this project.