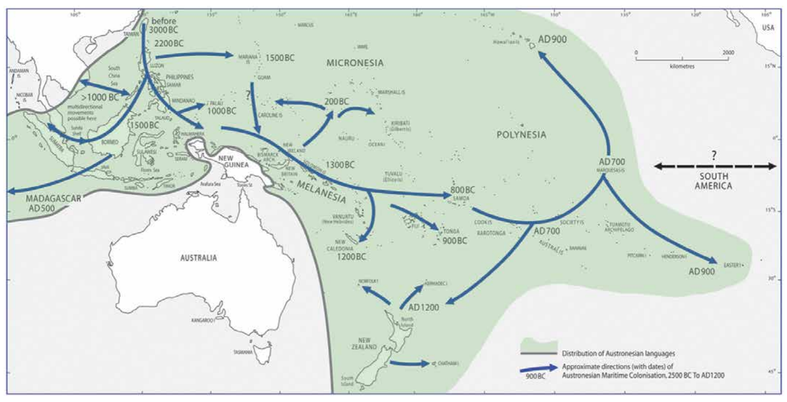
Science for Communities

Lisa Melia, Janet Stacey, Rachel Boyle and Bethany Forsythe.

Communication to Lisa Melia:
lisa.melia@esr.cri.nz

A whole mitochondrial genome MPS strategy for low-diversity populations



Benton M, Macartney-Coxson D, Eccles D, Griffiths L, Chambers G, et al. (2012) Complete Mitochondrial Genome sequencing Reveals Novel Haplotypes in a Polynesian Population. PLOS ONE 7(4): e35026. https://doi.org/10.1371/journal.pone.0035026 https://journals.plos.org/plosone/article?id+10.1371/journal.pone.0035026

New Zealand (NZ) is situated in the South Pacific. The main ethnic groups making up the population are categorised as self-declared Caucasians (mostly European or Middle Eastern including Iraqis and Iranians), Eastern Polynesians (including NZ Maori, Cook Islanders, Tokelauans, Hawaiians or Tahitians), Western Polynesians (Samoans, Tongans or Niueans) or peoples from Asia (mostly Chinese, Koreans, Japanese, or Filippinos). Genetic diversity within individual Pacific populations is low. About 5500 years ago (ya) Asian migrants from Taiwan became admixed with peoples from New Guinea and Melanesia prior to migration into Western Polynesia ~3000-3500 ya beginning in Fiji which has the highest overall genetic diversity. There is a gradual West-to-East decrease in overall diversity due to founder effects. Polynesian mitochondrial (mt) haplotypes include a 9bp deletion plus three temporally embedded substitutions known as the Polynesian motif. The 9bp intergenic deletion occurred ~60,000ya in Southeast Asia and subsequently three temporally embedded substitutions occurred at nucleotide positions (np) 16217, followed by np16261 ~6000ya in Taiwan and then np16247 in East Indonesia. The motif is now almost fixed in Polynesians. Migration proceeded to Tonga and Samoa ~4200-3000ya, into Eastern Polynesia about 2000-1000ya and finally from Tahiti to NZ via the Cook Islands ~750ya. The haplotype B4a1a1 is fixed in NZ Maori.

Sanger sequencing of the control region of mitochondrial DNA (mtDNA) has not been validated for casework use in NZ due to costs involved in having a purpose-built facility and a labour intensive workflow, together with reduced discrimination for Pacific Island and Maori subpopulations. ESR now has a sequencing facility and with the availability of third party analysis software for forensic applications, we are re-evaluating the costs and discrimination potential of whole genome mitochondrial sequencing using massively parallel sequencing (MPS).

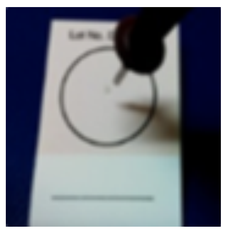
An evaluation was undertaken of a strategy designed to minimise errors from PCR and library preparation chemistries to enable reliable detection of low-level heteroplasmy for reference-type samples, including buccal samples on both Whatman® FTA® card and swabs with the registered mark R after Whatman and after FTA.

METHOD:

Briefly, samples were either extracted using DNA IQ™ (Promega) followed by real-time qPCR quantification of mtDNA (Kavlick et al, 2011, Sprouse et al, 2014) using HID Real-Time PCR Analysis Software version 1.2 (ThermoFisher Scientific), or direct amplification. Singleplex amplification of two overlapping amplicons was undertaken using PrimeSTAR® GXL DNA polymerase (TaKaRa Clontech), a hot start, long PCR polymerase recommended for high fidelity & mtDNA. Amplicons were prepared for sequencing using the PCR-free HyperPlus library preparation kit (Kapa) to reduce amplification bias & coverage dips. Libraries were sequenced on a MiSeq FGx™ (Illumina). Analysis was undertaken using GeneMarker®HTS software (SoftGenetics) blocking predictive health information sequences. Reads were aligned to the rCRS using the built in motif file to ensure correct forensic and phylogenetic nomenclature and default values were used for identity and soft clipping for the Illumina MiSeq. Variant filter settings used were: variant percentage ≥ 1%, variant allele coverage ≥ 40, total coverage ≥ 200, allele score difference ≤ 10 and allele balance ratio for SNPs (≤ 2.5) and indels (≤ 5.0) (Holland et al,2017, Riman et al,2017). Analysis reports provide mitotype data and haplotypes suitable for subsequent phylogenetic verification in EMPOP.

Direct PCR

- 1 x 2 mm punch for each of amplicons A and B (Fendt et al, 2009)
- multiple replicates processed

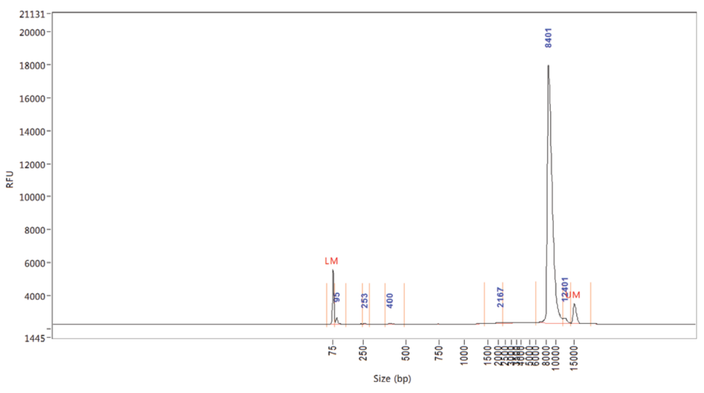


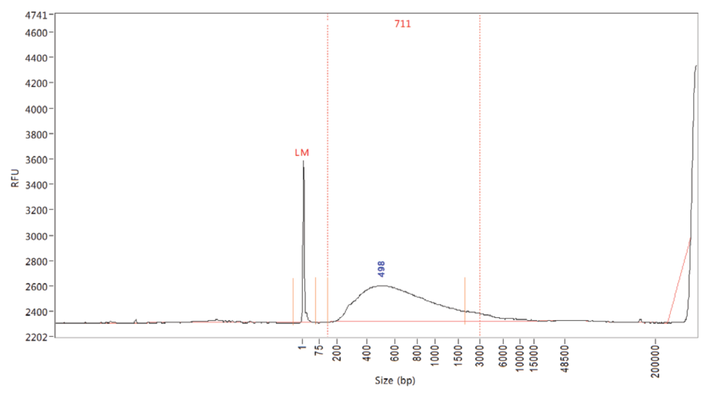
Long-range PCR: Takara PrimeSTAR® GLX

- Accurate long-range amplification (0.006% error rate cf Taq 0.04%) in GC- and AT-rich regions, robust in the presence of inhibitors
- modifications to manufacturer's protocol: increased dNTP, 2 x concentration of polymerase
- no inital denaturation
- half volume reaction (25ml)
- replicate amps
- quantified using Advanced Analytical Fragment Analyser dsDNA 75-15,000 bp reagent kit
- amplicons pooled in equal concentration, bead-purified using Bechman Coulter AMPpure XP and re-quantified.
- ***average pooled amplicons concentration from 4 replicates: 28.75 ng/μl***

Library prep: Kapa Hyperplus and dual-indexed Illumina adapter kit

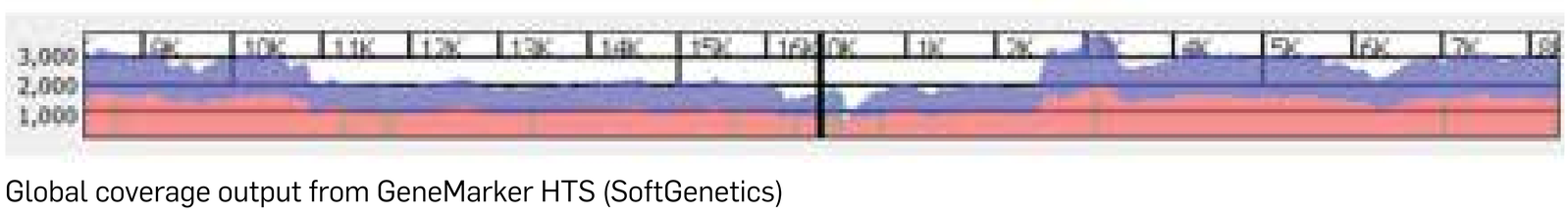
- coverage uniformity, lower coverage bias, PCR-free workflow
- half volume reactions (Ring et al, 2017)
- target input: 300ng
- libraries underwent two bead clean-ups and dimer removal was verified by fragment analysis and melt curves from real-time qPCR quantification
- libraries quantified using Kapa Library Quantification kit for Illumina platforms
- libraries were normalised to 4nM, then denatured and dilluted to 10pM and spiked with 5% 20pM PhiX
- sequenced on Illumina MiSeq Reagent Kit v3 using paired-end sequencing





Cluster density (k/mm²)	Clusters passing filter (%)	Total reads	Reads passing filter	% > Q30
575	96.68	14,030,000	13,560,000 (96%)	79.94

Sequencing run metrics



Global coverage output from GeneMarker HTS (SoftGenetics)

CONCLUSIONS:

This strategy shows promise as an option for jurisdictions with low-diversity subpopulations. Validation of this workflow is in progress. A population study will be progressed to characterise the discrimination potential for NZ subpopulations.

A full privacy impact assessment is being carried out to ensure compliance with NZ privacy legislation. It is likely that recommendations will include non-reporting of predictive health information, (including information in HV I and HV II) and blocking or masking of such sequencing information as soon as possible during analysis. Reports will not include mitotype sequencing information in line with our current reporting practice of non-inclusion of DNA profiling results.

References:

Fendt L, Zimmermann B, Daniaux M, Parson W. BMC Genomics 2009, 10: Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences.
Holland MM, Pack ED, McElhoe JA. For Sci Int: Genet 2017,28: Evaluation of GeneMarker® HTS for improved alignment of mtDNA MPS data, haplotype determination, and heteroplasmy assessment.
Kavlick MF, Lawrence HS, Merritt RT, Fisher C, Isenberg A, Robertson JM, Budowle B J For Sci 2011, 56: Quantification of human mitochondrial DNA using synthesized DNA standards.
Riman S, Kiesler KM, Borsuk LA, Vallone PM. For Sci Int: Genet 2017,29: Characterisation of NIST human mitochondrial DNA SRM-2392 and SRM-2392-I standard reference materials by next generation sequencing.
Ring JD, Sturk-Andreaggi K, Peck M, Marshall C. For Sci Int: Genet 2017, 29: A performance evaluation of Nextera XT and KAPA HyperPlus for rapid Illumina library preparation of long-range mitogenome amplicons.
Sprouse ML, Phillips NR, Kavlick BS, Roby RK. J Forensic Sci 2014, 59: Internal validation of human mitochondrial DNA quantification using real-time PCR.