Ancient mitochondrial DNA sequences from the First Australians revisited

Tim H. Heupink[a], S. Subramanian[a], Joanne L. Wright[a], Phillip Endicott[b,1], Michael C. Westaway[a], Leon Huynen[a], Walther Parson[cd] Craig D. Millar[e], Eske Willerslev[f2], David M. Lambert[a2]

[a] Environmental Futures Research Institute, Griffith University, 170 Kessels Road, Nathan, QLD 4111 Australia, [b] Department of Zoology, South Parks Road, Oxford University, Oxford, [c] Institute of Legal Medicine, Innsbruck Medical University, Austria, [d] Forensic Science Program, The Pennsylvania State University, University Park, Pennsylvania, USA, [e] Allan Wilson Centre for Molecular Ecology and Evolution, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland 1142, New Zealand, [f] Centre for GeoGenetics, University of Copenhagen, Denmark.

SI APPENDIX

Supplementary methods

Sequencing, base-calling and adapter trimming

All libraries, both non-capture and capture, were screened on a Bioanalyzer 2100 (Agilent) to ensure that the DNA length distributions did not show any significant artefacts from amplification (e.g. artificially long molecules due to serial binding). The library build blanks as well as the library builds for the extraction blanks were screened on the Bioanalyzer. The resulting libraries were 100bp single end sequenced on the HiSeq 2000 Sequencing System (Illumina). Sequences were base-called using CASAVA 1.8.2 (Illumina) and adapters were trimmed from the sequencing data using AdapterRemoval-1.5.4 (1) with the following options: a minimum 30bp length requirement after trimming and trimming terminal N's and low quality bases from the sequences using default parameters.

Determining levels of human DNA

First we determined the levels of human DNA by mapping the resulting reads to the complete human reference genome (GRCh37.p13, excluding alternative and patch scaffolds). Mapping was done using BWA 0.6.2-r126 (2) with the following options: seed disabled (3) and terminal low quality trimming (-q15). Unmapped and duplicate reads were removed from the resulting BAM file using Picard 1.68 (http://broadinstitute.github.io/picard/). The GenomeAnalysisToolKit-3.2-2 (GATK, 4) RealignerTargetCreator and IndelRealigner tools were used to improve alignment around insertions and deletions.

Determining authentic ancient DNA

Damage was estimated for each of the combined sample libraries using the program MapDamage (5), we averaged the proportions of 5' C>T and 3' G>A mutations on the terminal base to report damage. The GATK ReadLengthDistribution tool was used to count the number of occurrences of each read length.

Mitochondrial analysis

All reads were also mapped to only the revised Cambridge reference mitochondrial genome (6) and damage parameters were estimated where possible. The GATK UnifiedGenotyper with ploidy set to 1 was used to call a consensus vcf. The FastaAlternateReferenceMaker tool was then used to convert this to a fasta file based on positions with a Phred quality score of 30 or higher. The resulting sequence's regions that are prone to base call errors, i.e. the mononucleotide repeats around position 310, the ambiguous base at position 3107 and genuine deletions, were manually checked for errors. Coverage statistics were inferred using the DepthOfCoverage function in GATK. Haplogroups were inferred manually using mtDNA tree build 16 (7). Where multiple variants where observed the haplotypes and their level of presence were inferred by realigning the mapped reads to the RSRS (6, Table S4). The reads from WLH4 were sorted according to haplotype of origin and both datasets were subjected to the same methods of damage and read length inference, as described earlier.

Bioinformatics

Previously published complete Aboriginal Australian mitochondrial genomes and the sequence identified here for WLH4 were aligned using MAFFT v7.164b (8) using the default settings. The maximum likelihood tree was inferred using MEGA5.2 (9) with a GTR model and six gamma categories, ape and archaic human sequences were used to root the tree, but are not shown (Fig. 3, accession numbers are reported in Table S6). The mitochondrial sequence reported by Rasmussen et al. (10) was filtered for

calls with Phred quality scores of 30 or higher for this purpose. Haplogroups were inferred using HaploFind (11). The sequences from Adcock et al. (12) and those obtained in this study (trimmed to the same locus) were aligned using MAFFT (Table S7).

To estimate the age of the common most recent ancestor of WLH4 and modern Aboriginal Australians (MRCA) we collected all available complete mitochondrial genomes belonging to the haplogroup S2. We also included the mitogenomes of other representative Australian haplogroups along with African (L0 and L1 haplogroups) and Neanderthal outgroups. We constructed a Bayesian tree, using BEAST (13) using a HKY+G+I model and a relaxed log-normal molecular clock model. To calibrate the molecular clock we used the tip dates of Neanderthal (39,000 years) and WLH4 (1,600 years). We also used the age of human-Neanderthal split (320,000 – 620,000 years) to calibrate the root of the tree. A uniform prior distribution of 0 – 1E100 substitutions per site per year was used for the clock rate. A normal prior distribution with an mean of 470,000 years and a standard deviation was used in such a way that 95% of the prior calibration times fall between 320,000 – 620,000 years. These priors were based on the human-Neanderthal divergence estimated by a previous study using six Neanderthal mitochondrial genomes (14). The default priors were used for all other parameters. Based on the stratigraphic location of the burial of the skeletal remains, occlusal and task activity wear on the teeth, and lack of any mineralisation in the bone, the age of WLH4 can be estimated to be between 500 and 3,000 years old. Hence we used the mean value of 1,600 years for the calibration. However the age of the S2 haplogroup did not vary much when 500 and 3,000 years were used for the age of WLH4 (34.5 and 36.4 Kyr respectively).

PCR-based approach

We also obtained a quantity of the homogenised bone powder from WLH3 used in the original extractions (supplied by Gregory Adcock). The WLH3 bone was digested and purified using the silica binding method following the protocols of Brotherton et al. (15). A portion of the resulting extract (40uL) was made into a primary library, which was used in a targeted enrichment for human mtDNA. The resulting secondary library was quantified using quantitative realtime PCR before sequencing on an Ion Torrent personal genome machine.

After removal of the adaptors the sequences were aligned to the rCRS using the Geneious suite of software. A BAM file was exported and processed using PMD Tools. After filtering for damage, we did not observe any

sequences likely to originate from either endogenous or contaminating human mtDNA. This is consistent with attempts to amplify 50bp and 70bp human mtDNA fragments from the original extract using qPCR, which also returned no products.

We also investigated the original DNA extracts for WLH3 and KS8 used by Adcock et al. (12). After separation on a 3% agarose gel for 1 hour at 50 volts the products were excised under a low intensity UV light and purified using a Minelute column (Qiagen) according to the manufacturers instructions. The products were then cloned using the TopoTA system (Invitrogen) and multiple colonies amplified using T7 and M13R primers, and sequenced using an ABI capillary sequencer.

The SPEX assays followed the procedure detailed in (16), using the primer set targeting np 16224. Cloning was conducted using the TopoTA system (Invitrogen).

Phylogenetic re-examination of the HVR1 sequences

We re-examined the phylogenetic relationship and molecular divergences between ancient Australian mtDNA reported by Adcock et al. and those from contemporary Australians. For this purpose we obtained HVR1 region from 137 Australian Aborigines (Table S6). We also included HVR1 from LM3 (WLH3) and LM4 (WLH 4) reported by Adcock et al. and the ancient sequence obtained from WLH4 in this study. We used HVR1 sequences from Neanderthal, bonobo and chimpanzee as outgroups. This dataset is similar to that reported in Figure 1 of Adcock et al.. but with additional modern Aboriginal sequences. The HVR1 sequences were aligned using MUSCLE (17) with default settings. To obtain the best model of sequence evolution we used Modeltest implemented in the software MEGA5.2 (9). This analysis suggested the Tamura-Nei + Gamma model as the best, which was then used to construct the maximum likelihood tree using MEGA5.2, for visual clarity we excluded the Chimpanzee and Bonobo sequences (Figure S4).

We also estimated pairwise distances between ancient and modern First peoples (Figure S5). First we determined the extent of among-site variations based on a maximum likelihood method and obtained the gamma parameter. We then estimated the Tamura-Nei distance using the gamma value. Finally the pairwise Tamura-Nei distances were multiplied by the length of the HVR1 to obtain the number of differences. Since this method accounts for multiple hits as well as base compositional bias the pairwise differences obtained are more accurate.

In order to investigate the ability of the locus used by Adcock et al. (12) to correctly infer phylogenetic relationships we aligned the complete mitochondrial genomes reported by Ingman et al. (18) and additional ape and archaic human sequences (accession numbers in Table S6) using MAFFT (8). The maximum likelihood tree was inferred using RaxML (19, with GTRCAT model and gorilla forced as outgroup, Fig. S5A). The same approach was used for the same dataset trimmed to amplicon reported by Adcock et al. (positions 16047-16399 on the rCRS, Fig. S5B).

Willandra Lakes Community and Elder approval

This research has been conducted in partnership with the Willandra Council of Elders. Before applying to Australian Research Council funding to support the research, we consulted extensively with the Elders regarding the nature and extent of the proposed study. We emphasised that, because the skeletal remains is, in most cases, very old, we could not be confident of being able to recover any DNA sequences from them. We have met with them on two occasions to discuss the work. At their meeting on Monday 1st March 2010, the Elders of the Barkindji and Ngiyampaa People unanimously resolved to support this application and to agree to DNA sampling of the Willandra collection of sub-fossil human remains. These were, at the time, held at the Australian National University. Later in 2013 the Muthi Muthi People returned to the management structure of the Willandra Lakes World Heritage Area. We met with the Muthi Muthi Elders and provided them with an overview of the research results. This was also discussed in a plain language report providing an overview of the results. At this time we also met with the Barkindji and Ngiyampaa to provide an overview of the results. Finally a draft of this paper was sent to all Traditional Owners for comment.

Willandra Lakes Region World Heritage Area

5th May, 2010

Dr Michael Westaway
Department of Archaeology
GPO Box 2100
Adelaide SA 5001

Dear Michael

At the Two Traditional Tribal Groups of the Willandra Lakes Elders Council meeting held on Monday 1st March, 2010 approval was granted by the Elders for your request for a research proposal, to examine skeletal DNA from Willandra Lakes Region Skeletal remains. We would also like to accept your offer to be involved during the process of this proposal when taking samples from remains that are currently housed at the Australian National University.

A minimum of two Tribal Elders will need to be present. It is expected that you will cover their travel and accommodation costs. This office can assist in organising Elders attendance and in approving access to the collections in Canberra.

Note that the two Traditional Tribal Groups are the Paakantji and the Ngyiampaa and that acknowledgement should be given to these groups in any published material.

I look forward to hearing from you.

Regards,

Richard Mintern
Willandra Lakes World Heritage Area
Executive Officer
Ph: 03 5021 8908

Phone: (03) 5021 8908
Fax: (03) 5022 2037

Table S1 - Genomic mapping results for each library

Library	Number of reads	After trimming	Number of mapped reads	%	Unique reads	%	Duplication (%)
WLH3.a Lib1	25,101,862	22,550,698	215,389	0.96	154,199	0.68	28.4
WLH3.a Lib2	25,156,974	22,414,695	620,112	2.77	173,996	0.78	71.9
WLH3.a Lib3	119,339,411	106,286,647	7,729,959	7.27	5,627,144	5.29	27.2
WLH3.a Lib4	29,651,240	28,054,447	14,666,540	52.28	6,690,697	23.85	54.4
WLH3.b Lib1	6,308,912	4,959,856	43,575	0.88	42,938	0.87	1.5
WLH3.b Lib2	18,642,556	16,026,385	129,316	0.81	120,450	0.75	6.9
WLH3.b Lib3	163,285,552	139,106,711	597,545	0.43	557,290	0.40	6.7
WLH3.b Lib4	15,195,584	13,990,868	2,827,564	20.21	1,530,158	10.94	45.9
WLH4 Lib1	1,106,086,943	1,072,269,973	27,892,736	2.60	5,039,684	0.47	81.9
WLH4 Lib2	1,248,113,984	1,127,661,246	34,514,023	3.06	23,366,949	2.07	32.3
WLH4 Lib3	1,072,627,373	1,029,105,777	34,376,116	3.34	15,494,830	1.51	54.9
WLH15.a Lib1	33,425,874	30,734,878	3,547	0.01	3,491	0.01	1.6
WLH15.b Lib1	17,315,925	16,140,276	25,344	0.16	22,725	0.14	10.3
WLH55 Lib1	11,890,710	11,386,223	1,181	0.01	1,151	0.01	2.5
WLH55 Lib2	15,306,560	14,440,116	2,311	0.02	2,281	0.02	1.3
WLH3IT Lib1	182,128	49,144	207	0.42	144	0.29	30.4

Table S2 - Genomic and mitochondrial capture statistics

	Before capture			After capture			Enrichment	
Reference	**Library**	**Mapped Reads (%)**	**Unique Reads (%)**	**Library**	**Mapped Reads (%)**	**Unique Reads (%)**	**Mapped Reads (X)**	**Unique Reads (X)**
Genomic	WLH3.a Lib3	7.27	5.29	WLH3.a Lib4	52.82	23.85	7.2	4.5
Genomic	WLH3.b Lib3	0.43	0.40	WLH3.b Lib4	20.21	10.94	47.0	27.4
Mitochondrial	WLH3.a Lib3	0.0053	0.0035	WLH3.a Lib4	0.679	0.0339	127.6	9.7
Mitochondrial	WLH3.b Lib3	0.0003	0.0003	WLH3.b Lib4	0.283	0.0326	874.4	109.7

Table S3 - Read lengths and molecular damage observed for nuclear and mitochondrial sequences

	Human genome (GRCh37)			Mitochondrial genome (rCRS)		
Sample	**Number of Unique reads**	**Average read Length**	**Terminal Base C>T**	**Number of Unique reads**	**Average read Length**	**Terminal Base C>T**
WLH3.a	10,479,485	79.1	0.01	9,749	79.3	0.01
WLH3.b	2,130,967	78.5	0.01	4,701	77.6	0.01
WLH4	43,293,836	67.8	0.07	27,623	75.4	0.13
WLH15	26,216	55.7	0.01	6	69.7	NA
WLH55	3,432	60.5	0.00	1	94.0	NA
WLH3IT	122	51.4	NA	29	49.4	NA

Table S4 - Highly variable nucleotide sites in the mapped mitochondrial sequences

WLH3.a

Position	**Reference base**	**Alternate base**	**Reference count**	**Alternate count**	**Alternate percent**	**Associated haplotype**
56	A	AG	8	36	82%	
73	G	A	7	50	88%	H ancestral
309	C	T	27	4	13%	

2706	G	A	8	33	80%	H ancestral
3010	G	A	45	4	8%	H1
3093	C	A,T	35	3	8%	
3110	C	A,T	30	3	9%	
3197	T	C	37	7	16%	U5a'b
6221	T	C	41	5	11%	
6253	T	C	13	32	71%	H15a1
6691	GAAAAAAA	GAAAAAA	36	3	8%	
7028	T	C	5	40	89%	H ancestral
7379	G	A	32	5	14%	
7400	C	T,A	33	3	8%	
9477	G	A	51	9	15%	U5a'b
11410	T	C	16	25	61%	H15a1
11467	A	G	38	9	19%	U5a'b
11719	A	G	5	43	90%	H ancestral
12308	A	G	47	6	11%	U5a'b
12372	G	A,T	40	8	17%	U5a'b
13617	T	C	42	7	14%	U5a'b
13759	G	A,T	44	4	8%	
14766	T	C	6	45	88%	H ancestral
14793	A	G	50	8	14%	U5a'b
14953	C	T	9	27	75%	H15a1
15301	G	A	34	3	8%	

WLH3.b

Position	Reference base	Alternate base	Reference count	Alternate count	Alternate percent	Associated haplotype
195	C	T	2	13	87%	
309	C	T	17	3	15%	
310	T	C	13	5	28%	
3010	G	A	22	5	19%	H1
4218	T	C	11	10	48%	
6776	T	C	20	5	20%	H3
7444	G	A	4	11	73%	H40b
7621	T	C	10	8	44%	H40b
7678	T	C	11	13	54%	H40b
8063	T	C	22	6	21%	
15378	T	G,A	13	2	13%	

WLH4

Position	Reference base	Alternate base	Reference count	Alternate count	Alternate percent	Associated haplotype
72	T	C	87	32	27%	V3c
73	G	A	86	35	29%	V3c
152	C	T	89	16	15%	

2380	C	T	18	96	84%	S2
3438	G	A	21	93	82%	S2
4580	G	A	103	29	22%	V3c
5261	G	A	23	119	84%	
6167	T	C	24	99	80%	S2
7283	T	C	27	84	76%	
8281	C	A	22	79	78%	
8404	T	C	20	93	82%	S2
8537	A	G	32	88	73%	
9077	T	C	103	19	16%	
11719	A	G,T	104	28	21%	V3c
12705	T	C	90	16	15%	V3c
12810	A	G	110	23	17%	V3c
14766	T	C	89	13	13%	V3c
15236	A	G	24	103	81%	
15904	C	T	91	16	15%	V3c
16072	C	T	21	93	82%	
16179	C	T	13	80	86%	
16184	C	T	13	81	86%	
16216	A	G	98	13	12%	V3c
16298	T	C	89	16	15%	V3c

Table S5 - Details for library construction.

Remains	PCR protocol	Primary Cycles	Secondary Cycles	Capture	Post-Capture Cycle
WLH3a library 1	AccuPrime	12	8	-	-
WLH3a library 2	AccuPrime	12	8	-	-
WLH3a library 3	Kapa HiFi Uracil+	10	8	Before capture library	-
WLH3a library 4	Kapa HiFi Uracil+	10	8	After capture library	16
WLH3b library 1	AccuPrime	12	0	-	-
WLH3b library 2	AccuPrime	12	0	-	-
WLH3b library 3	Kapa HiFi Uracil+	10	4	Before capture library	-
WLH3b library 4	Kapa HiFi Uracil+	10	4	After capture library	18
WLH4 library 1	AccuPrime	12	8	-	-
WLH4 library 2	AccuPrime	12	0	-	-
WLH4 library 3	Platinum Taq HiFi	12	12	-	-
WLH15a library 1	AccuPrime	12	0	-	-
WLH15b library 1	AccuPrime	12	0	-	-
WLH55 library 1	AccuPrime	12	0	-	-
WLH55 library 2	AccuPrime	12	0	-	-
WLH3IT library 1	IonTorrent	-	-	-	-

Table S6 – Sequence Details

Table S6A - Details of the sequences shown in Figure 2

Number	Origin	Haplogroup	Accession Number	Note
not shown	Bonobo		NC001644	
not shown	Chimp		NC001643	
not shown	Denisovan		FR695060	
not shown	Denisovan		FN673705	
not shown	Gorilla		NC011120	
not shown	Neanderthal		KJ533544	Vi33.17
not shown	Neanderthal		KJ533545	Vi33.19
not shown	Neanderthal		FM865407	Feldhofer1
not shown	Neanderthal		FM865408	Feldhofer2
not shown	Neanderthal		FM865409	ElSidron1253
not shown	Neanderthal		FM865410	Vindija33.25
not shown	Neanderthal		FM865411	Mezmaiskaya1
1	AU	S1a	DQ404440	
2	AU/Desert	S1a	DQ404441	
3	AU	N14	DQ112753	
4	AU	S1	AF346963	
5	AU	S	DQ112751	
6	AU/Kalumburu	S5	EF495220	
7	AU	S4	AY289062	
8	AU	S3	AY289066	
9	AU	S3	AY289067	
10	AU	S2	AF346965	
11	AU	S2	AY289060	
12	AU	S2	AY289051	
13	AU	S2	AY289061	
14	AU/Kalumburu	N13	EF495214	
15	AU	O	AY289059	
16	AU	O1	DQ404447	
17	AU/Kalgoorlie	O1a	NA	Rasmussen et al. (2011)
18	AU	O1a	AY289056	
19	AU	O1a	AY289058	
20	AU	R12	DQ112752	
21	AU	P3a	AY289052	
22	AU	P3a	AY289065	
23	AU/ArnhemLand	P3b1	EF061153	
24	AU	P5	AY289063	
25	AU	P7	AY289054	
26	AU	P6	AY289053	
27	AU	P6	AY289055	
28	AU/Riverine	P8	DQ404446	
29	AU	P4b	AY289064	
30	AU	P4b1	AY289057	
31	AU/Riverine	P4b1	DQ404444	
32	AU	K1d	DQ112750	
33	AU/Kalumburu	Q2b	EF495218	
34	AU/Kalumburu	M15	EF495219	
35	AU/Kalumburu	M14	EF495222	

36	AU	M42a	DQ404443	
37	AU	M42a	DQ404445	
38	AU/Desert	M42a	DQ404442	
39	AU	M42a	DQ112754	
40	AU	M42a	AF346964	
41	AU	M42a	DQ112755	

Table S6B - Details of the sequences shown in Figure S3

Modern Aboriginal Australians	Neanderthal	Ancient Australians
AF039317	FM865407	AF328745 - LM3 Adcock et.al (2001)
AF039318	FM865408	AF328746 – LM4 Adcock et.al (2001)
AF039319		
AF039320		
AF039321		
AF039322		
AF039323		
AF039324		
AF039325		
AF039326		
AF039327		
AF039328		
AF039329		
AF039330		
AF039331		
AF039332		
AF039333		
AF039334		
AF039335		
AF039336		
AF039337		
AF039338		
AF039339		
AF039340		
AF039341		
AF039342		
AF039343		
AF039344		
AF039345		
AF039346		
AF039347		
AF039348		
AF039349		
AF039350		
AF039351		
AF039352		
AF039353		
AF039354		
AF039355		

AF039356		
AF039357		
AF039358		
AF039359		
AF039360		
AF039361		
AF039362		
AF039363		
AF176126		
AF176128		
AF176129		
AF176131		
AF176132		
AF176133		
AF176134		
AF176135		
AF176136		
AF176137		
AF176138		
AF176139		
AF176140		
AF176141		
AF176142		
AF176143		
AF176144		
AF176156		
AF176157		
AF176158		
AF176159		
AF176160		
AF176164		
AF176165		
AF176166		
AF176168		
AF176169		
AF176170		
AF176171		
AF176172		
AF176173		
AF176175		
AF176176		
AF176177		
AF176181		
AF176183		
AF176185		
AF176186		
AF176187		
AF176188		

AF176189		
AF176190		
AF176191		
AF176192		
AF176193		
AF176194		
AF176195		
AF176200		
AF176201		
AF176202		
AF346963		
AF346964		
AY289051		
AY289052		
AY289053		
AY289054		
AY289055		
AY289056		
AY289057		
AY289058		
AY289059		
AY289060		
AY289061		
AY289062		
AY289063		
AY289064		
AY289065		
AY289066		
AY289067		
DQ404440		
DQ404441		
DQ404442		
DQ404443		
DQ404444		
DQ404445		
DQ404446		
DQ404447		
EF495214		
EF495218		
EF495219		
EF495220		
EF495222		
JN226143		
JN226144		
JN226145		
S80333		
U37730		
U37731		

U37732		
U37733		

Table S6C - Details of the sequences shown in Figure S5A

Number	Origin	Haplogroup	Accession number	Note
1	Korean	B4a1b1a	AF346993	
2	Samoan	B4a1a1o	AF347007	
3	Piman	B2a5	AF347001	
4	Uzbek	B4c2a	AF347011	
5	Italian	U5b3a1a	AF346988	
6	Tatar	H5	AF346974	
7	Dutch	H5a1a	AF346975	
8	French	H1r	AF346981	
9	Saami	V7a1	AF347006	
10	English	HV0	AF346978	
11	Georgian	T2e	AF346982	
12	German	J1c4	AF346983	
13	China	F4a1b	AF346973	
14	PNG	P1d1	AF347005	
15	PNG	P1	AF347002	
16	PNG	P1	AF347004	
17	Australian	S2	AF346965	
18	Australian	S1	AF346963	
19	Chukchi	A2b1	AF346971	
20	Buriat	C4b3	AF346970	
21	Khirgiz	C4b1	AF346991	
22	Evenki	C4a2	AF346979	
23	Warao	C1d1	AF347012	
24	Warao	C1d1	AF347013	
25	Australian	M42a	AF346964	
26	PNG	Q1b	AF347003	
27	Inuit	D2a1b	AF347010	
28	Japanese	D4b2b1	AF346989	
29	Japanese	D4a1	AF346990	
30	Guarani	D1a2	AF346984	
31	China	M9a1a1a	AF346972	
32	Indian	G3b1	AF346966	
33	Yoruba	L3e2b1a1	AF347015	
34	Lisongo	L3e2b7	AF346994	
35	Bamileke	L3e3b2	AF346967	
36	Ewondo	L3e1e1	AF346980	
37	Yoruba	L3d1a1a	AF347014	
38	Mkamba	L3h1a2a1	AF347000	
39	Effik	L2a1a2	AF346977	
40	Effik	L2a1i1	AF346976	
41	Mandenka	L2c3a	AF346995	
42	Mbenzele	L1c1a1a1a	AF346997	

43	Biaka	L1c1a1a1a	AF346968	
44	Biaka	L1c1a2b	AF346969	
45	Mbenzele	L1c1a2b	AF346996	
46	Ibo	L1c1d1	AF346987	
47	Kikuyu	L1c2a1a	AF346992	
48	Ibo	L1b1a3	AF346986	
49	Mbuti	L0a2b	AF346999	
50	Mbuti	L0a2b	AF346998	
51	Hausa	L0a1a2	AF346985	
52	San	L0k1a1c	AF347009	
53	San	L0k1a1a	AF347008	
54	Neanderthal		FM865410	Vindija33.25
55	Neanderthal		FM865407	Feldhofer1
56	Neanderthal		KJ533544	Vi33.17
57	Neanderthal		KJ533545	Vi33.19
58	Neanderthal		FM865409	ElSidron1253
59	Neanderthal		FM865408	Feldhofer2
60	Neanderthal		FM865411	Mezmaiskaya1
61	Denisovan		FN673705	
62	Denisovan		FR695060	
63	Bonobo		NC001644	
64	Chimp		NC001643	
65	Gorilla		NC011120	

Table S6D – Details of the sequences shown in Figure S5B

Number	Origin	Haplogroup	Accession number	Note
60	Neanderthal		FM865411	Mezmaiskaya1
59	Neanderthal		FM865408	Feldhofer2
56	Neanderthal		KJ533544	Vi33.17
57	Neanderthal		KJ533545	Vi33.19
55	Neanderthal		FM865407	Feldhofer1
54	Neanderthal		FM865410	Vindija33.25
58	Neanderthal		FM865409	ElSidron1253
64	Chimp		NC001643	
63	Bonobo		NC001644	
50	Mbuti	L0a2b	AF346998	
49	Mbuti	L0a2b	AF346999	
51	Hausa	L0a1a2	AF346985	
53	San	L0k1a1a	AF347008	
52	San	L0k1a1c	AF347009	
62	Denisovan		FR695060	
61	Denisovan		FN673705	
42	Mbenzele	L1c1a1a1a	AF346997	
43	Biaka	L1c1a1a1a	AF346968	
45	Mbenzele	L1c1a2b	AF346996	
44	Biaka	L1c1a2b	AF346969	
46	Ibo	L1c1d1	AF346987	

47	Kikuyu	L1c2a1a	AF346992	
48	Ibo	L1b1a3	AF346986	
38	Mkamba	L3h1a2a1	AF347000	
25	Australian	M42a	AF346964	
17	Australian	S2	AF346965	
7	Dutch	H5a1a	AF346975	
27	Inuit	D2a1b	AF347010	
19	Chukchi	A2b1	AF346971	
28	Japanese	D4b2b1	AF346989	
30	Guarani	D1a2	AF346984	
32	Indian	G3b1	AF346966	
29	Japanese	D4a1	AF346990	
31	China	M9a1a1a	AF346972	
40	Effik	L2a1i1	AF346976	
39	Effik	L2a1a2	AF346977	
41	Mandenka	L2c3a	AF346995	
1	Korean	B4a1b1a	AF346993	
2	Samoan	B4a1a1o	AF347007	
3	Piman	B2a5	AF347001	
4	Uzbek	B4c2a	AF347011	
33	Yoruba	L3e2b1a1	AF347015	
34	Lisongo	L3e2b7	AF346994	
9	Saami	V7a1	AF347006	
10	English	HV0	AF346978	
13	China	F4a1b	AF346973	
12	German	J1c4	AF346983	
11	Georgian	T2e	AF346982	
6	Tatar	H5	AF346974	
8	French	H1r	AF346981	
15	PNG	P1	AF347002	
14	PNG	P1d1	AF347005	
16	PNG	P1	AF347004	
5	Italian	U5b3a1a	AF346988	
18	Australian	S1	AF346963	
37	Yoruba	L3d1a1a	AF347014	
26	PNG	Q1b	AF347003	
35	Bamileke	L3e3b2	AF346967	
36	Ewondo	L3e1e1	AF346980	
24	Warao	C1d1	AF347013	
23	Warao	C1d1	AF347012	
22	Evenki	C4a2	AF346979	
21	Khirgiz	C4b1	AF346991	
20	Buriat	C4b3	AF346970	
65	Gorilla		NC011120	

Table S7. Alignment of polymorphic nucleotide positions for sequences recovered in this study from the Willandra Lakes material and that from Kow Swamp.

	16093	16126	16179	16184	16201	16216	16223	16224	16230	16253	16259	16263	16264	16278	16284	16290	16298	16301	16311	16355	16356	16387	16397
rCRS (Reference)	T	T	C	C	C	A	C	T	A	A	C	T	C	C	A	C	T	C	T	C	T	A	T
AT[1]	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
GJA[1]	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	**Y**	.	.	.	.
Contaminant[1]	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
LM3[1]	.	.	.	.	.	.	T	.	G	.	.	C	T	T	.	T	.	T	.	T	C	G	.
WLH3 PCR 38/38 clones[2]						.	.	C	.	.	.	.	.	.	.	.	.	.	C				
WLH3 SPEX 24/24[2]				.	.	.	.	.															
WLH3a(H15a1) HISEQ[3]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
WLH3a(U5a) HISEQ[3]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
WLH3a(H1) HISEQ[3]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
WLH3b(H40b) HISEQ[3]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
WLH3b(H1) HISEQ[3]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
WLH3b(H3) HISEQ[3]	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
KS8[1]	.	.	.	.	.	.	T	.	G	.	.	.	.	T	G	.	.	.	C	.	.	.	C
KS8 PCR 18/23 clones[2]						.	T	.	.	.	.	.	.	T	.	.	.	.	.				
KS8 PCR 5/23 clones[2]						.	T	.	.	.	A	.	.	T	.	.	.	.	.				
KS8 SPEX 13/17 clones[2]					.	.	.	.															
KS8 SPEX 4/17 clones[2]					.	.	T	.															

LM4[1]	.	.	.	.	T	.	.	.	.	G	.	.	.	.	.	.	.	.	C	.	.	.	.						
WLH4(S2) HISEQ[3]	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.						
WLH4(V3c) HISEQ[3]	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.						

Nucleotide positions on the revised Cambridge reference mitochondrial genome are given above each column. [1] indicates the sequence was observed by Adcock et al. 2001, [2] by the laboratory in Oxford, UK and [3] by the laboratory in Brisbane, Australia. AT and GJA are the sequences of Alan Thorne and Gregory Adcock themselves, the contaminant represents a sequence regularly observed and reported by Adock et al 2001. A "." indicates the sequence its nucleotide matches that of the reference sequence, absence of a character indicates the essay did not recover or target the nucleotide. The bold "Y" indicates a base difference (cytosine or thymine) for GJA's Genbank and Table 1 sequences. Polymorphisms that occurred only once in the PCR and SPEX clones have been disregarded for this table.

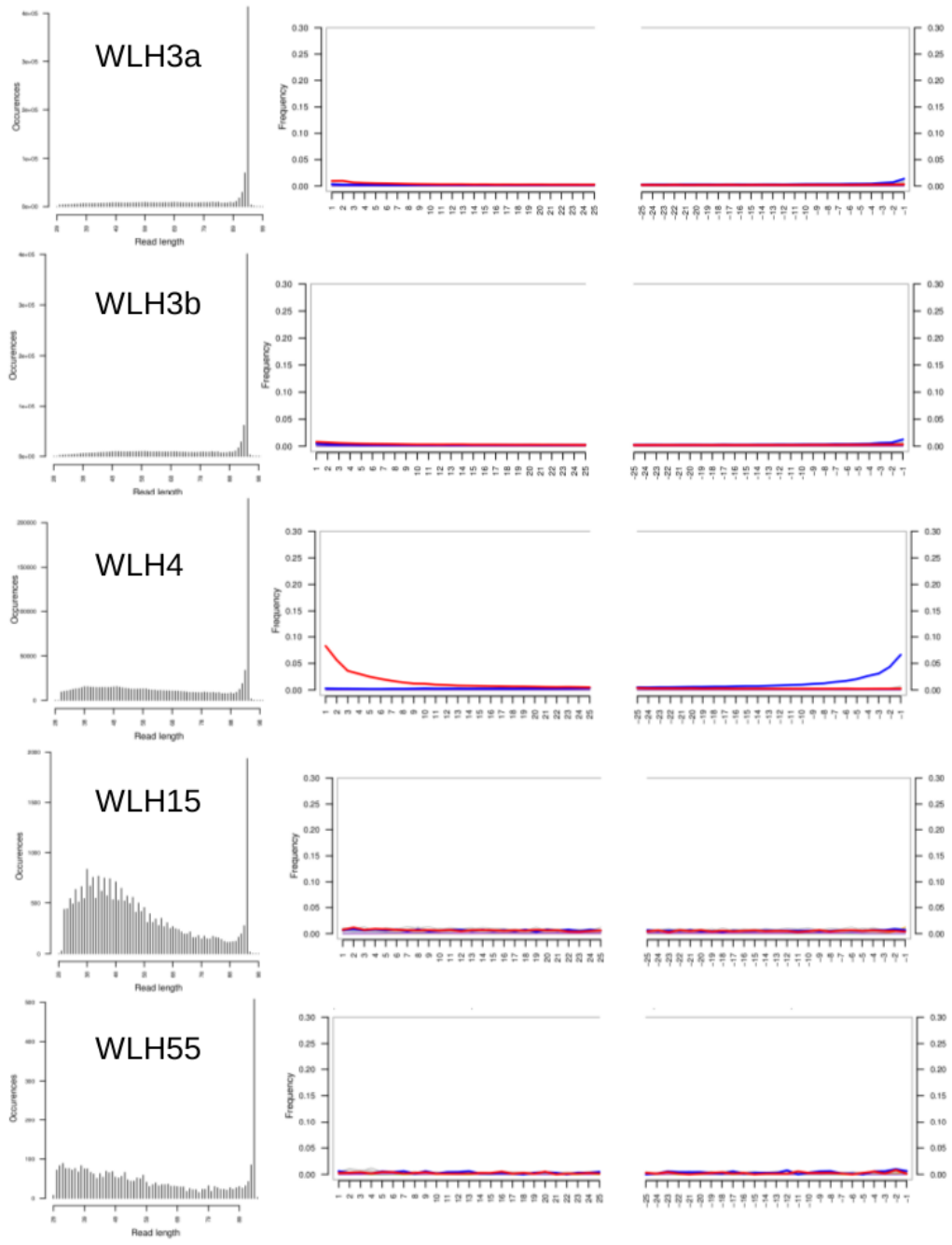

Figure S1. Read length (A) and damage distributions (B) for genomic data.

Molecular damage as represented by 3' guanine to adenine substitutions (red) and 5' cytosine to thymine substitutions (blue) in recovered DNA sequences for Willandra Lakes samples WLH3 (samples a and b), WLH4, WLH15 and WLH55.

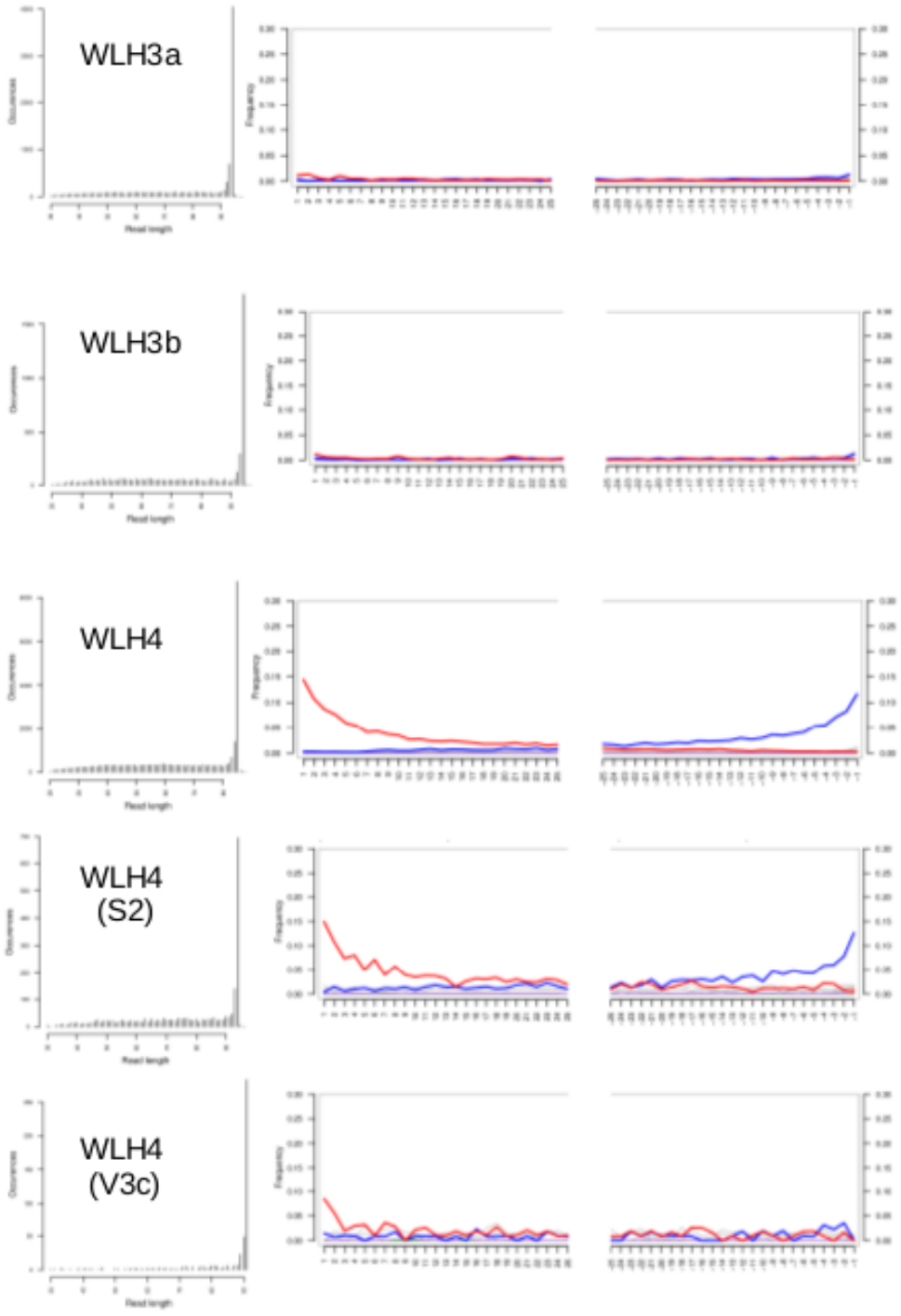

Figure S2 Read length and damage distributions for mitochondrial data. Damage is represented by observed 3' guanine to adenine substitutions (blue) and 5' cytosine to thymine substitutions (red).

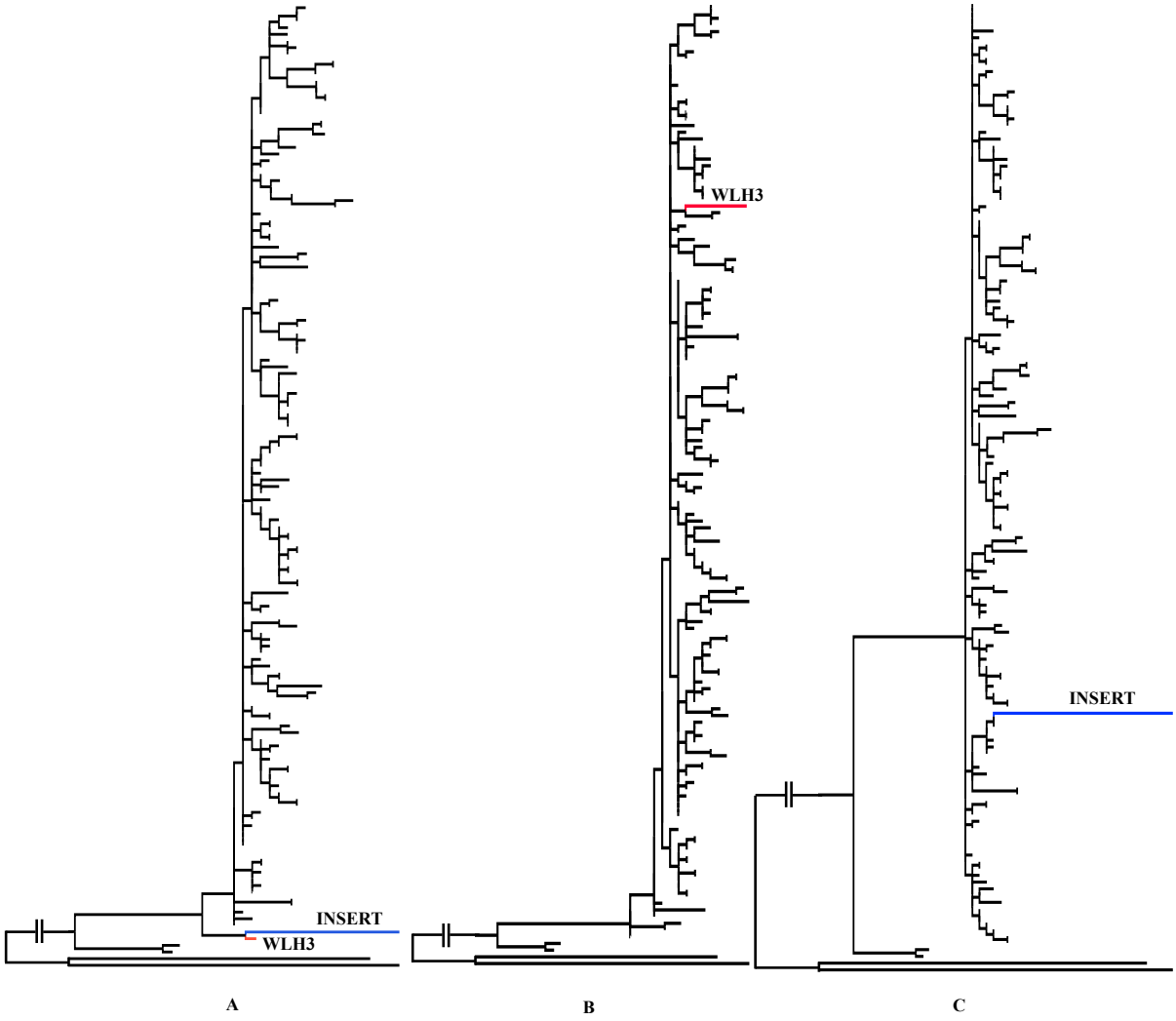

Figure S3. Maximum likelihood trees of Aboriginal Australian based on mitochondrial HVR1 sequences. The trees were constructed using ancient and 137 modern HVR1 sequences along with a nuclear insert of HVR1, Chimpanzee, Bonobo and Neanderthal sequences. (A) Includes both WLH3 (LM3) and the nuclear insert sequences (B) Only WLH3 and (C) Only the insert sequence. This figure shows that WLH3 falls outside the rest of the Australian Aboriginal HVR sequences only when WLH3 and the nuclear insert sequence is included. Note that the tree is rooted using Chimpanzee and Bonobo sequences, which are not shown in the figure.

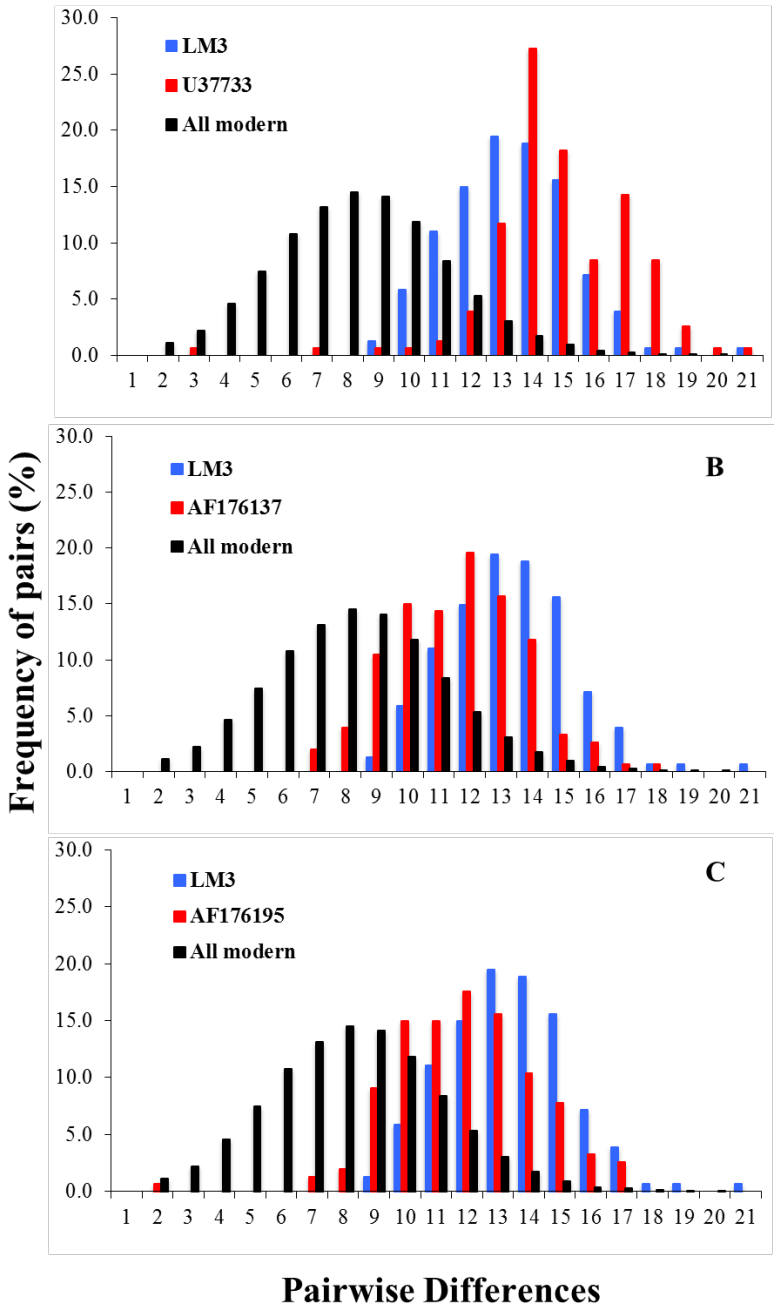

Figure S4. Pairwise distances for Australian Aborigines

Pairwise distances estimated within contemporary Australian Aborigines (Black), between LM3 (WLH3) and modern Australians (Blue) and between fast evolving and other remaining modern Australians (Red).

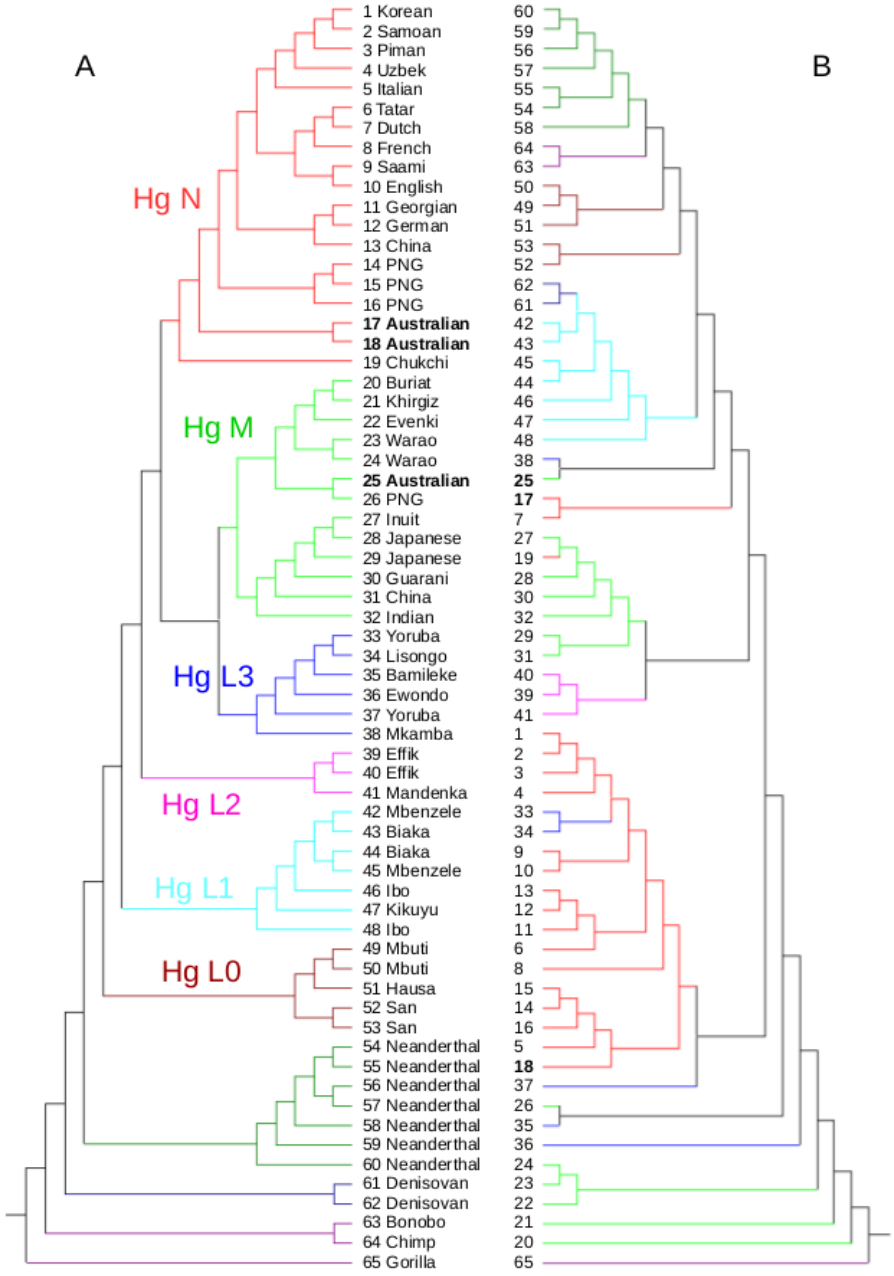

Figure S5. Comparison of Maximum Likelihood trees.
Only the topology is shown for clarity. The data in Figure S5A are complete mitochondrial genomes. Figure S5B shows relationships among sequences for the same dataset but trimmed to the amplicon reported by Adcock et al. (2001). The braches are colour coded as follows: Apes (purple), Denisovans (dark blue), Neanderthals (dark green), major human haplogroups: L0 (dark red), L1 (turquoise), L2 (pink), L3 (blue), M* (green), N* (red). Gorilla was forced as the outgroup. Details of the sequences can be found in Table S6.

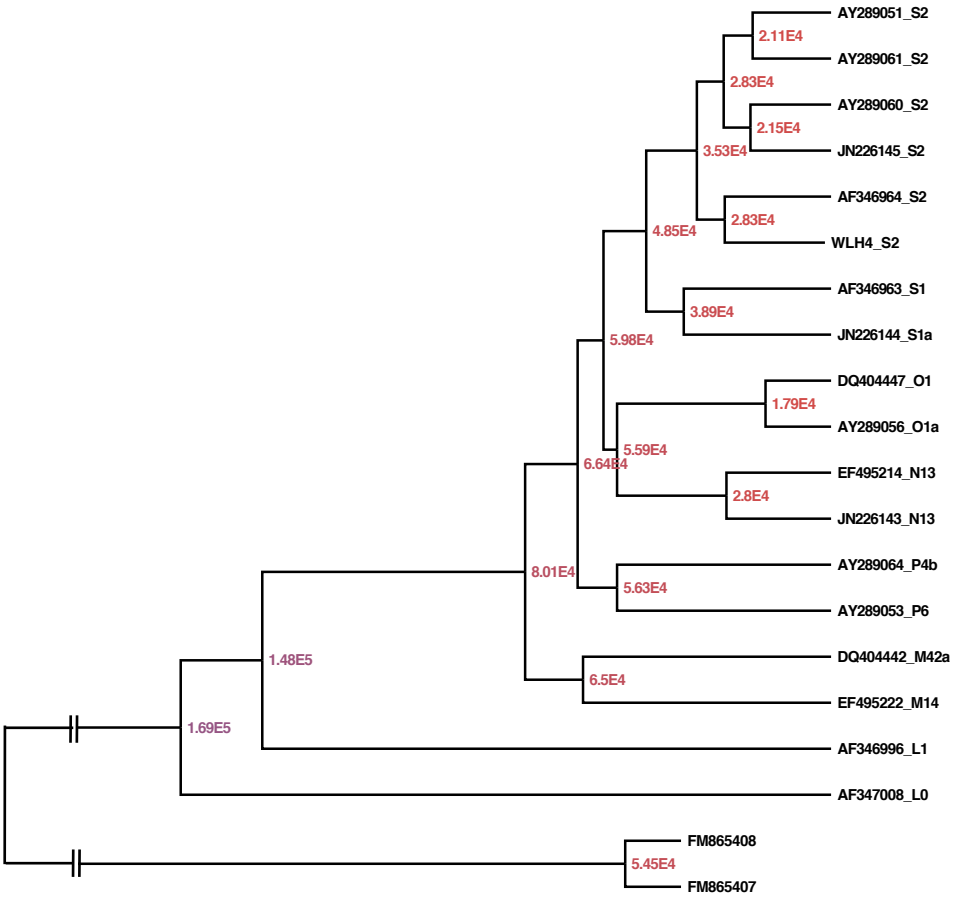

Figure S6. Bayesian time tree of Australian haplogroups

The time tree of Australian S2 haplogroup along with other Australian haplogroups, African and Neanderthal outgroups. The complete mitochondrial genomes were used to construct a Bayesian tree, using BEAST. The tree was calibrated using the tip dates of Neanderthal (39,000 years) and WLH4 (1,600 years) and the age of human-Neanderthal split (320,000 – 620,000 years) was also used to calibrate the root of the tree (see methods).

References

1. Lindgreen S (2012) AdapterRemoval: easy cleaning of next-generation sequencing reads. BMC Research Notes 5(1):337.

2. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26(5):589-595.

3. Schubert M, et al. (2012) Improving ancient DNA read mapping against modern reference genomes. BMC genomics 13(1):178.

4. McKenna A, et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research 20(9):1297-1303.

5. Jónsson H, Ginolhac A, Schubert M, Johnson PL, Orlando L (2013) mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics btt193.

6. Andrews RM, et al. (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature genetics 23(2):147-147.

7. Van Oven M, Kayser M (2009) Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human mutation 30(2):E386-E394.

8. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution 30(4):772-780.

9. Tamura K, et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular biology and evolution 28(10):2731-2739.

10. Rasmussen M, et al. (2011) An Aboriginal Australian genome reveals separate human dispersals into Asia. Science 334(6052):94-98.

11. Vianello D, et al. (2013) HAPLOFIND: A New Method for High-Throughput mtDNA Haplogroup Assignment. Human mutation 34(9):1189-1194.

12. Adcock GJ, et al. (2001) Mitochondrial DNA sequences in ancient Australians: implications for modern human origins. Proc Natl Acad Sci USA 98(2):537-542.

13. Drummond AJ, Suchard MA, Xie D & Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7 Molecular Biology And Evolution 29: 1969-1973.

14. Briggs, AW et.al (2009) Targeted Retrieval and Analysis of Five Neandertal mtDNA Genomes. Science 325, 318-321.

15. Brotherton P, et al. (2013) Neolithic mitochondrial H genomes and the genetic origins of Europeans. Nature Communications 4:1764.

16. Brotherton P, Sanchez JJ, Cooper A, Endicott P (2009) Preferential access to genetic information from endogenous hominin ancient DNA and accurate quantitative SNP-typing via SPEX. Nucleic Acids Research 38(2):e7.

17. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic acids research 32(5):1792-1797.

18. Ingman M, Kaessmann H, Pääbo S, Gyllensten U (2000) Mitochondrial genome variation

and the origin of modern humans. Nature 408(6813):708-713.

19. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22(21):2688-2690.