**Supporting Information Corrected October 27, 2014**

## *Supplementary Materials and Methods*

**Sample collection.**

Thirty-seven ancient chicken bones were collected for analysis, comprising: eight ancient chicken bones from archaeological sites at Paluki and Anatoloa in Niue; 11 ancient Hawaiian chicken bones from an excavation at Makauwahi Cave on Kauai, Hawaii collected by DB; and 18 Rapa Nui chicken bones excavated from deposits at Anakena collected by TH. The 18 Rapa Nui bones include the six samples previously analyzed by Storey *et al.* (1) (Table S1). One hundred and twenty four modern feather samples were also examined to investigate recent phylogeographic patterns. These included 107 modern feathers from ISEA and Remote Oceania collected by GL and KD in 2008 and 2009: 28 from the Santa Cruz Islands, 31 from the Solomon Islands, 13 from Papua New Guinea, 10 from Indonesia, 23 from the Philippines and two from Vietnam. An additional 17 naturally shed modern feather samples were collected from the Marquesas (French Polynesia, n=6) by TH, and Kokee, Kauai (Hawaii, n=11) by TH/DB. Details on the locations of these modern samples can be found in Table S1 and are shown in Fig. 1.

**Ancient DNA Extraction, Amplification, and Sequencing.**

The samples were extracted, amplified, and sequenced in specialist ancient DNA (aDNA) laboratories at the Australian Centre for Ancient DNA (ACAD) in Adelaide, South Australia, according to a range of strict protocols and including controls (2). Ancient bone samples (n=37) were extracted and PCR experiments set up in the physically remote ACAD ancient laboratory, whereas the feathers (n=17) were extracted and PCR experiments set up in the physically remote ACAD pre-PCR clean-room laboratory. Independent external replication of the ACAD9068 (PAQANA011) ancient sample was performed in a dedicated aDNA lab in the Archaeology Department at Durham University following strict laboratory procedures (2).

**ACAD ancient bone extractions.**

Each chicken bone was ground to fine powder in a Mikrodismembrator (5000 rpm, for 10 seconds). Approximately 70 mg of bone powder was decalcified concurrently with protein digestion by incubation at 55 °C overnight in 1mL of extraction buffer (consisting of 0.4725 M EDTA (pH=8.0), 0.2 % sodium dodecyl sulphate (SDS), and 0.7 $mg.ml^{-1}$ Proteinase K). After digestion, samples were centrifuged at 10,000 rpm for 5 mins and the supernatant was transferred to an Amicon ultra-4 (Millipore), which was centrifuged at 4000 xg until only 100 µL supernatant remained. The supernatant was washed with 1 mL molecular grade water and centrifuged again (at 4000 xg until only 100 µl remained). An equal volume of ATL buffer (Qiagen DNeasy kit) was then added, mixed, and the supernatant removed to a 2 mL screw-cap tube. The supernatant was incubated for 10–60 mins at room temperature on a rotary mixer after the addition of an equal volume of AL buffer (Qiagen DNeasy kit) and 0.02 $\mu g.\mu l^{-1}$ of carrier RNA. After the incubation, an equal volume of ethanol (100 %) was added, and then the total volume was transferred to a Qiagen DNeasy spin column where it was incubated at room temperature for 10–60 mins. The extraction then followed the Qiagen DNeasy kit instructions, with the following exceptions at the elution stage: 100–150 µL of warmed AE buffer was added and then incubated at room temperature for 10–30 mins, before being centrifuged at 8,000 rpm for 1 min, this step was repeated to finish with 200–300 µL of total volume.

**ACAD PCR amplification and sequencing of ancient samples.**

A 330 base pair (bp) segment of the mtDNA CR was amplified and sequenced from each specimen in short overlapping fragments (Table S10, Fig. S13), which is necessary to ensure amplification of the short damaged fragments of ancient DNA samples. PCRs were set up using 25 µL volumes containing a final concentration of 1 U Platinum Taq DNA Polymerase High Fidelity (Invitrogen), 1 x PCR Buffer (Platinum, Invitrogen), 3 mM MgSO4, 200 µM each dNTP, 2 $mg.ml^{-1}$ rabbit serum albumin (Sigma), 1 µM forward and reverse primers and 2-3 µl of template DNA. PCR reactions were performed on a Corbett Research Palm Cycler using the following cycling conditions: 94 °C for 2 min, 55 cycles of 94 °C for 30 s, 55 °C for 30 s, 68 °C for 30 s, and a final extension of 10 min at 68 °C. Amplifications of extraction and PCR controls were performed in all experiments to monitor contamination. PCR products were separated by electrophoresis on a 3.5 % agarose gel. Successful PCR products (10 µl) were purified using 0.8 µl of EXOSAP (Fermentas) at final concentration of 0.38 U/µl Exonuclease I, and 0.05 U/µl Shrimp Alkaline Phosphotase, and thermal cycled at 37 °C for 30 mins, 80 °C for 15 mins, and 15 °C for 3 mins on a Corbett Research Palm Cycler. The forward and reverse complements of each fragment were sequenced from the same PCR reaction using the same primers as for the PCR, and Big Dye Terminator v3.1 cycle-sequencing chemistry, followed by vacuum clean up on a Multiscreen®$_{384}$ SEQ plate (Millipore). The sequencing run was conducted on an ABI 3130XC capillary sequencer.

Primers GG144F/GG387R and GG316F/GG586R (1) were used initially to amplify a portion of the mitochondrial (mtDNA) control region but as the PCR products amplified from these primers (fragment 1

and 2) are 250bp and 305bp respectively; additional primers were designed to cover the same range of mtDNA control region. Primer GG144F was paired with A1781 (187bp as fragment 3) and A1780 was paired with GG387R (151bp as fragment 4) to cover the equivalent DNA sequence as fragment 1 but in two overlapping fragments (Table S10). Primers A1958 and A1959 (192bp as fragment 5) were used to cover the balance of the mtDNA CR under study for the ancient samples. The use of this alternative primer set meant that a sequence gap was introduced in some ancient sequences equivalent to the primer binding region (Fig. S13). Further trimming to the sequence length shared across all chicken specimens resulted in a final sequence length of 201bp.

**Durham Bone extractions as replication for PAQANA011.**

DNA extraction of the replicate ancient chicken bone fragment PAQANA011 was performed in a dedicated aDNA lab in the Archaeology department at Durham University following strict laboratory procedures as per commonly used guidelines (2). All equipment and work surfaces were cleaned before and after each use with a dilute solution of bleach (10 %) followed by ethanol (99 %). The ancient chicken bone (~0.05 g) was pulverized in a Micro-dismembrator, digested in 0.425 M EDTA, 0.05 % SDS, 0.05 M Tris-HCI and 0.333 $mg.ml^{-1}$ proteinase K and incubated overnight on a rotary mixer at 50 $^{o}$C until fully dissolved. 2 ml of solution was then concentrated in a Millipore Amicon Ultra-4 30 KDa MWCO to a final volume of 100 µl. The concentrated extract was purified using the QIAquick PCR Purification Kit following manufacturers recommendations, except that the final elution step was performed twice to produce a final volume of 100 µl. A negative extraction control was performed alongside the ancient bone sample.

**Durham PCR amplification and sequencing of ancient samples.**

PCRs were setup in 25 µl reactions using 1.25 U Taq GOLD (Applied Biosystems), 1 x Gold buffer (Applied Biosystems), 2.5 mM $MgCl_2$, 0.5 $\mu g.\mu l^{-1}$ bovine serum albumin (BSA), 200 µM of each dNTP, 0.8 µM of each forward and reverse primers, and 2-5 µl of aDNA extract. We used PCR primers (5'-3') GG144F and GG387R; GG316F, and GG586R (1). One PCR negative control was included for every three aDNA template PCR tubes. We ran a total of 22 PCRs with aDNA template, eight PCR negative controls and two PCR negative extraction control. Neither the PCR negative controls nor the negative extraction control produced bands (PCR product) when analyzed by gel-electrophoresis. PCR cycling conditions were 95$^{o}$C for 5 min, 50 cycles of 94 °C for 45 sec, 54 °C for 45 sec and 72 °C for 45 sec, followed by 72 °C for 10 min. PCR products were stored at -20 °C. Sanger sequencing on the Applied Biosystems 3730 DNA Analyser was performed at the DNA sequencing service in the School of Biological and Biomedical Sciences at Durham University.

**Modern DNA Extraction, Amplification, and Sequencing.**

**ACAD modern feather extractions.**

Approximately 5 mm of each feather tip was rehydrated overnight with 1 ml phosphate buffered saline (PBS) on a rotary mixer at room temperature. On day 2, the supernatant was removed, the feather tip was macerated using a clean scalpel blade, and the sample was digested in 440 µl of digestion buffer (comprising ATL buffer (Qiagen DNeasy kit) with 1.8 $mg.ml^{-1}$ Proteinase K, and 90 mM Dithiothreitol) overnight at 55 °C on a rotary mixer. After digestion, 400 µL of AL buffer (Qiagen DNeasy kit) and 0.02 $\mu g.\mu l^{-1}$ of carrier RNA was added and incubated at room temperature on a rotary mixer for 10–30 mins, after which 400 µL of 100 % ethanol was added. The supernatant (650 µl) was incubated on a Qiagen DNeasy spin column for 10–30 mins before being centrifuged at 8000rpm for 1 min. This incubation was then repeated until all of the supernatant had been centrifuged through the column. The feather extraction protocol then followed that of the bone extraction procedure above.

**ACAD PCR amplification and sequencing of modern feather samples.**

PCR amplifications and sequencing of the 2 overlapping fragments were performed as per the ancient bone samples (see above).

**Durham modern feather extraction**

At Durham University, modern feathers from ISEA and Near Oceania were extracted in a pre-PCR clean room after Cooper & Poinar (2), using a protocol designed by Pfeiffer *et al.* (3) alongside the QIAquick PCR purification Kit (QIAGEN Ltd, UK). The tip of each feather was sampled (approximately 1cm cut into smaller fragments) and digested in 340µl extraction buffer containing 100mM Tris-HCl, pH8, 100mM NaCl, 3mM $CaCl_2$, 2% SDS (w/v), 40mM DTT and 250µg/ml proteinase K following the protocol by Pfeiffer *et al.* (3). The samples were incubated overnight at 56°C on a rotary mixer. Following digestion, the samples were purified using the QIAquick PCR purification Kit (QIAGEN Ltd, UK) following the manufacturers' instructions. An extraction control was used for every run of seven samples. The quantity of DNA present within each extract was measured using the Quant-iT HS Assay Kit (Invitrogen) used with the

Qubit fluorometer following the manufacturers' instructions.

**Durham PCR amplification and sequencing of modern feather samples.**

The amplification of a 201bp fragment of the CR (a subset of the 330bp amplified from the ancient samples) was undertaken through PCR in a physically separated clean laboratory. The forward primer GG144F and the reverse primer GG387R (see Table S10) were used to amplify this 201bp fragment (excluding primers). The PCR amplifications were performed in a 25µl reaction mix containing 1µl of extract, 0.96x PCR Gold Buffer, 2.4mM $MgCl_2$, 1.2U *Taq*, 0.24mM dNTP and 0.96µM of each primer. The PCR thermal cycling reactions consisted of 90s initial denaturation step at 94°C, followed by 35 cycles of 30s denaturation at 94°C, 30s annealing at 54°C, 30s extension at 72°C then a 10 minute final extension step at 72°C. The PCR products were visualized on a 0.5x agarose gel. Sequencing was performed on an ABI 3730 sequencer in the DNA-dedicated laboratory of the School of Biological and Biomedical Sciences.

**Cloning of PAQANA011 at ACAD.**

The PCR products generated from bone sample PAQANA011 were cloned using Stratagene and/or Topo cloning kits using manufacturers instructions (after an A-tailing reaction). The A-tailing reaction consisted of a 20 µl volume reaction containing 0.125 U HotMaster Taq, 2.5 µM dATP, 10x HotMaster buffer, 17 µl cleaned PCR products. The Buffer, dATP's and Taq were activated at 94 °C for 2 mins prior to addition of the PCR products then a further incubation at 72 °C for 10 mins. The A-tailed PCR products were then cleaned up using an Isopropanol precipitation and resuspended in 10 µl of PCR grade water. Sanger sequencing of the cloned PCR products were performed according to the procedures outlined above.

**Phylogenetic inference**

**WMG:** To determine the robustness of the current phylogenetic framework used for chicken research, the 61 WMG sequences from that study were downloaded and aligned using the Muscle algorithm in Geneious v5.6 (4). PartitionFinder v1.0.1 (5) was used to identify the number of preferred partitions and their substitution model (CR with HKY plus Gamma; codon 1, codon 2 and tRNA with HKY; and codon 3 with GTR). MrBayes v3.2 was used to generate a phylogenetic tree using four runs of four independent chains of 100 million iterations, less 25% as burnin (6). Tests for convergence to stationarity were performed by analyzing the standard deviation of split frequencies (< 0.01). RaxML v7.0.4 was used to generate a maximum likelihood tree with the same partitions as above, with bootstrapping performed via 100 iterations followed by an optimized maximum likelihood search (7).

To establish the level of phylogenetic concordance between topologies produced by WMGs versus the highly variable 201bp of the CR, the WMG data was split into two subsets, the 201bp fragment of the CR and the WMG excluding all of the CR. Each subset was rerun for the PartitionFinder and MrBayes analyses separately (*i.e.* the CR was run separately from the WMG data minus the CR), using the same parameters as above except only 2 million iterations were required to obtain a standard deviation of less than 0.01 for the four chains.

**mtDNA CR:** In addition to the 144 CR sequences generated in this study, we downloaded 1226 worldwide mtDNA CR chicken sequences from Genbank (1, 8-17) to establish the geographic distribution for each chicken haplogroup (n=1370). Although additional CR sequences have since been uploaded to Genbank (total chicken CR sequences are currently >3000), overall haplogroup designations are not changed with the inclusion of additional sequences (18). To allow direct comparisons of the CR haplotypes, the 1370 chicken sequences were aligned and trimmed to the 201bp common to our 144 newly generated sequences (referred to as 'full CR dataset'), with any indels removed. The 201bp hypervariable fragment is a useful region for reconstructing recent evolutionary events when DNA template length is a constraint (19, 20), such as in ancient DNA studies. For ease and clarity, the 1370 CR sequences were collapsed to unique haplotypes using Collapse v1.2 with manual adjustments where missing data caused short sequences to be considered different haplotypes, resulting in 274 unique haplotypes (H001-H274, see Dataset S6; referred to as 'unique CR haplotype dataset'). The haplogroup of each of our 144 newly generated sequences was established by comparison to sequences of known haplogroup designation from Liu *et al.* (13) (see Dataset S6). The phylogenetic robustness of the full 330bp length (both fragment 1 & 2) was investigated using PhyML (21) to establish that inclusion of additional length sequences did not change the haplogroup designation of the new sequences (Fig. S14), with ModelGenerator (22) used to establish the model of best fit. We also explored the unique CR haplotype dataset in SplitsTree4 (23), using the NeighborNet algorithm, and found that the data appeared not to be tree-like, probably due to saturation and substitution rate heterogeneity (18). As the majority of the new 144 CR sequences were identified as haplogroup D, a Median Joining Network (using Network v4.6; 24) was also generated for just the D haplogroup. DNAsp was used to generate the

input file for the Network program. As DNAsp does not allow ambiguous bases and as these ambiguous bases were assumed to reflect sequencing errors, each ambiguous base was modified to reflect the more common of the possible bases within the haplogroup. Default weights were used in Network. To examine the discrepancies between the composition and phylogeographic distribution of haplogroups reported by Storey *et al*. (1, 25, 26) and those generated in this study, we tested the likelihood of detecting the reported proportions. Tests of statistical significance were performed using the binom.test command and probability distribution graphs were created using the dbinom command (Fig. S9), in the R 'stats' package (27). A linear regression plot (Fig. S10) was also generated to visualize the correlation between occurrence of the characteristic 4 CR SNPs of the Polynesian chicken and longitude using the standard plotting function in R (27). Population genetic and differentiation statistics were estimated in Arlequin v3.5 (28) for each population.

**Bayesian Serial Simcoal (BayeSSC) simulations**

Bayesian coalescent simulations (using Bayesian Serial Simcoal – BayeSSC v1.0; 29) were used to model eight possible scenarios of chicken colonization of the New World via either 1) Polynesia or 2) Europe. Low level migration between populations was 1) permitted or 2) not permitted, and two separate datasets were examined: 1) only containing haplogroup D ancient samples (representing authenticated Polynesian chicken signals); and 2) containing all putative ancient haplotypes (ancient samples from haplogroups B, D, E; 1, 25, 26; this paper ). In order to test between the different migration routes in BayeSSC, we modeled the same uniform priors for modern population deme size and population growth for each of the migration scenarios to maintain similar demographic parameters.

All eight of the South American migration simulations were performed using common uniform priors on modern effective population sizes (MSEA: 10,000-2,000,000; ISEA: 10,000-1,000,000; Europe: 10,000-1,000,000; South America: 1,000-1,000,000; and Pacific: 1,000-1,000,000), with the total panmixia model having a uniform prior with a slightly lower minimum and slightly higher maximum (10,000-10,000,000). The uniform prior on the growth rate since the last migration event (which differs for each model – see Figure S11) was also common across all eight migration scenarios (growth rate of -0.00001, which equates to 0.001% per generation). Although the generation time of free-ranging domestic chickens is not known, we have estimated a generation time of a year. We considered this appropriate as we were attempting to model early historic chicken populations, which would have had relatively short life spans and low fecundity due to their value as a food source of both meat and eggs. The samples included in the BayeSSC simulations and the migration matrices used are provided in Tables S2-3 and S4-5, respectively.

To explore likely demographic histories for chickens in western Polynesia, we also used BayeSSC to simulate alternate migration route hypotheses for comparison with the observed phylogeographic patterns within the Pacific. Sequences in the 201bp CR dataset from the Pacific and ISEA that had location details (n=177) were used to model five possible scenarios of migration routes through western Polynesia, Micronesia and eastern Polynesia (see Fig. S12): a total panmixia model; two models that describe the colonization of Micronesia but with no onward link to Polynesia (one from the Philippines-Micronesia [P-M; arrow 2A in Fig. 1] and the other from New Guinea-Micronesia [NG-M; arrow 3]); and two models that describe Micronesia as a stopping point in an onward route to Polynesia (one from the Philippines-Micronesia-West Polynesia [P-M-WP; arrows 2A and 2B] and the other from New Guinea-Micronesia-West Polynesia [NG-M-WP; arrows 3 and 2B]). Note that the alternate scenario of migration from Micronesia to New Guinea was not tested. The Pacific migration scenarios also had common uniform priors on modern effective population sizes (Philippines: 10,000-2,000,000; PNG: 10,000-2,000,000; Micronesia: 1,000-1,000,000; Melanesia: 1,000-1,000,000; Western Polynesia: 1,000-1,000,000; Eastern Polynesia: 1,000-1,000,000), and a common uniform prior on growth rate since the last migration event at 750 BP (growth rate of -0.00001, which equates to 0.001% per generation). The samples included in the Pacific BayeSSC simulations and the migration matrices used are provided in Tables S7 and S8, respectively

## *Supplementary Information*

**Ancient Pacific sample (PAQANA011)**

Repeated amplifications and Sanger sequencing of Storey *et al.*'s PAQANA011 sample (1) placed it within the D haplogroup (Dataset S3), however it also highlighted 10 type 2 transitions (C-to-T or G-to-A) across the 12 amplicons. This type of transition is commonly observed in aDNA because of post-mortem template damage, with the hydrolytic loss of amino-groups from cytosine converting the base to uracil, which DNA polymerases read as a thymine base (30). As these internal PCR replications confirmed the discrepancy between our extraction (ACAD9068) and Storey *et al.*'s (1) published sequence (EF535246) for

bone sample PAQANA011, we had it independently replicated by another aDNA laboratory at Durham University (Dataset S3).

A subsample of the PAQANA011 bone was sent to Durham University, where it was extracted and three PCR amplifications were performed for each of fragments 1 and 2 of the mtDNA CR (as only fragment 1 is diagnostic to haplogroup level and below, it is this fragment that is compared to the Liu *et al.* (13) dataset in the discussion). Two different haplotypes were detected across the three amplicons of fragment 1 for PAQANA011. Two amplicons matched each other and fell within clade D. The third amplicon matched Haplotype A35, which is found in chickens from China and Japan (13), as well as two other Genbank samples: AM746039 (14) and AB263973, both of which are commercial breeds. Liu *et al.* (13) found almost 95% of domestic chickens to belong to clades A, B, C, E, F, and G, which suggests the presence of clade A (this study) and clade E (1) amplicons from this sample may reflect lab consumable/reagent contamination by modern domestic chicken DNA. Although fragment 2 is not as phylogeographically informative as fragment 1, amplification of this fragment did reinforce the highly damaged/degraded nature of this particular sample, which may have allowed modern chicken DNA (at low levels in lab reagents/consumables) to occasionally outcompete the endogenous DNA. Across the three amplicons of fragment 2, seven randomly distributed C-to-T transitions were found to differentiate the sequences from our extract of this sample (ACAD9068; see Dataset S3). Although the damaged sites tended to be within the longer amplicons, this was not always the case.

Cloning of the PCR amplicons from both labs (ACAD and Durham University) was undertaken at ACAD to confirm the Sanger sequencing results and to establish whether the C=>T transitions were due to post-mortem damage. At the base pairs where C=>T transitions were initially detected, few differences were detected between the clones, which suggests that each amplicon was formed by amplification from a single damaged template (Dataset S4). However, at least four matching (non-damaged) amplicons were retrieved for both fragments 1 and 2, so that a consensus sequence could be generated (Dataset S4).

**Shrimp DNase experiment to test ACAD3890 and ACAD9060 samples**

From the 24 ancient samples successfully amplifying DNA, two samples (ACAD3890 from Niue, and ACAD9060 from Rapa Nui) yielded haplotypes other than from clade D. ACAD3890 matched Liu *et al.'s* haplotype 'A34', which is found in only one modern sample from Xinjiang, China (13). ACAD3890 had poor amplification/sequencing success (n=1/31), with only one amplicon (102bp) amplifying and sequencing successfully (using primers A1780 and GG387R). The second sample, ACAD9060, matched E01 found commonly worldwide (China, n=19; India, n=10; Sri Lanka, n=20; Japan, n=27; Iran, n=3; Turkmenistan, n=3; UK, n=2; Europe, n=34; Chile, n=25, Kenya, n=58) (11, 13, 15, 16). This sample also did not amplify often (n=8/14), with the E01 haplotype only occurring once (all other PCR amplicons could not be successfully sequenced). Accordingly, contamination by modern chicken DNA was suspected as the source of both the A34 and E01 haplotypes, however due to the stringent aDNA procedures in place at the ACAD, there is limited opportunity for modern chicken DNA to enter the lab. The possibility of contamination in the laboratory consumables/reagents was tested by the addition of Shrimp DNase to three sets of subsequent PCR reactions for all 24 samples. Shrimp DNase is an endonuclease that cleaves phosphodiester bonds in double stranded DNA. It is often used to treat PCR master mixes prior to the addition of extracted DNA in order to break down contaminating modern DNA in PCR reagents.

No DNA was successfully amplified after Shrimp DNase treatment of PCR's for the ACAD3890 sample, while Shrimp DNase treatment of PCR's for ACAD9060 sporadically gave haplotype D sequence across a variety of fragment sizes (210bp, 190bp, 129bp, and 90bp), plus one sequence (116bp) that could not be assigned to any haplogroup – it had 7 mismatches from its closest BLAST matches (94% identity). Both of these results suggest PCR reagent contamination by modern chicken DNA was the likely source of the original A34 and E01 haplotypes. When ACAD9060 did give a D haplotype, it matched the D haplotype from the other ancient Pacific samples for fragment 1, however fragment 2 could not be amplified so this sample was excluded from further analyses. The possibility of laboratory consumable/reagent contamination needs to be discussed more in aDNA studies, especially when the use of a simple PCR additive, such as Shrimp DNase, can rule out one source of possible contamination (i.e. PCR reagents; 31). This additive (or similar) is essential for aDNA studies of commensal or domesticated species, where DNA from modern populations of the same species may permeate factories where lab consumables and/or reagents are produced (32).

Although the overall percentage of E haplotypes (15/53, 28%) detected in ancient Polynesian chicken samples is higher than the nominal 5% contamination rate of modern domestic species found in lab consumables (32), low levels of preserved endogenous DNA may allow any contaminating modern chicken DNA to outcompete them in PCR reactions (unless an endonuclease such as Shrimp DNase is used to remove contaminating DNA in PCR lab reagents). Previously, some of the Storey *et al*. co-authors have

reported low amplification success from Mele Havea (Tonga) and Paluki (Niue) (33) — and in our analyses of Paluki material we identified a non-D haplotype to be contamination from PCR reagents (Dataset S1).

**Haplogroup E vs. haplogroup D**

More broadly, haplogroup D has been found to closely follow the distribution of cockfighting in India, Indonesia, China and Japan (13). Many Polynesian societies have traditionally supported cockfighting (called 'faatitoraamoa' in Tahitian; 34), for example Tahitians had many songs and religious traditions (including 'Ruaifaatoa', the god of cockfighting) connected to faatitoraamoa (35). In contrast many of the other haplogroups are ubiquitous worldwide, potentially as a result of early historic dispersal with European colonialists (*e.g.* haplogroups A, B, and E) and are therefore likely to be phylogeographically uninformative and the predominant contamination of laboratory consumables.

To investigate the conflicting results obtained here versus those previously reported by Storey *et al.* (1, 25, 26), we calculated the probability of detecting the reported proportions of D and E haplogroups given the different datasets. Tests of statistical significance were performed using the binom.test command and probability distribution graphs were created using the dbinom command in the R 'stats' package (Fig. S9; 1, 25, 26). It is possible that if haplogroup E was present in low frequencies amongst ancient Pacific chickens (*e.g.* 10%) we did not detect it within our 22 ancient samples simply due to stochastic sampling effects (P-value = 0.098). However, if E was actually present at only 10% of the ancient Pacific chickens then it is highly unlikely that Storey *et al.* would also have detected 15/31 ancient Pacific chickens as having haplogroup E sequences (P-value = $6.9 \times 10^{-9}$).

**BayeSSC coalescent simulations for testing South American link**

A European source of South American chickens was also the more likely scenario when only haplogroup D sequences were considered to represent authentic Polynesian chickens (*i.e.* simulations using all modern chicken data, but only haplogroup D ancient samples). This is perhaps not surprising as the only ancient haplogroup D sample from South America (from early historic Peru) does not share a haplotype with any ancient population in Eastern Polynesia. In fact, the most geographically proximate Pacific populations sharing the early historic Peruvian haplotype (H033) are Vanuatu, Santa Cruz and the Solomon Islands, and all ISEA populations contain this common haplotype. The movement of chickens between South America and the Philippines via the Manila galleon trade in the 1500s (20, 25) may provide a possible explanation for the presence of this common ISEA haplotype in early historic Peru.

**BayeSSC coalescent simulations for testing migration routes to Micronesia**

The Bayesian simulations suggest the most likely scenario of those tested involved movement of chickens between Micronesia and the Bismarck Sea at a relatively early date (although post-human arrival in Micronesia) but with little interaction with chickens further eastward. A link between the long-distance trade and communication network of Yap (in Micronesia) and the Bismarck region has previously been postulated by Kirch based on linguistics, with further archaeological evidence suggesting Fais was originally settled from Yap about 100 AD (36, 37). An early migration route linking the Bismarcks with islands in Micronesia, via the proposed trade hub of the 'Yapese Empire', is therefore not surprising. A newly-discovered Lapita migration route along the southern coast of PNG at 2500 BP, further highlights the extensive nature of early Polynesian networks (38).

Only one ancient specimen has likely influenced the Bismarck/Micronesian link (a Fais sample dated to 600 ± 40 BP with H260; 25). However, the coalescent has still been able to reconstruct the true history even though the same haplotype is present in modern chickens from the Solomon and Santa Cruz Islands (*i.e.* the coalescent hasn't been overwhelmed by a shared haplotype). Rather, it is the entire Micronesian chicken population (including ancient H260 and modern H032, H224, and H225 haplotypes) that has contributed to the coalescent reconstruction of the migration model. As the timing of this Bismarck-Micronesia link is based purely on coalescent simulations using an inferred mutation rate, the proposed Bismarck-Micronesia migration route is not necessarily temporally robust.

**Issues with previous radiocarbon dates**

The three pre-Columbian dates reported in Storey *et al.* (NZA 26115, NZA28271 and NZA28272; 1) were performed without ultrafiltration of high molecular weight collagen, or the removal of exogenous organic matter through XAD-2 purification. It is also notable that the oldest sample (NZA26115) was a very small sample characterized by a low collagen yield, and lacks analytical data. Previous work has shown that neither C:N ratios, nor d13C:d15N ratios are indicators of reliable radiocarbon dates (39). Given the importance of this result it would be highly desirable that these (and other) specimens were re-dated using

the most advanced methodologies available. These would include d13C measurements of individual amino acids (40) to properly examine the potential for dietary marine carbon to produce an erroneously older (pre-Columbian) date.

***Table S1. Pacific samples used in this study with corresponding haplogroups discussed in the text.***

| | *Locations* | | *# of samples* | | | *No, of successful samples by haplogroup* | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | *Island* | *Site name* | *# attempted* | *# successful* | *# contaminated* | *D* | *E* | *A* | *B* | *I* | *References* |
| ANCIENT | Rapa Nui | Anakena | 18 | 13 | 1[#] | 13 | 1[#] | | | | This study |
| | Hawaii | Makauwahi | 11 | 7 | 0 | 7 | | | | | |
| | Niue | Anatoloa | 2 | 2 | 0 | 2 | | | | | |
| | | Paluki | 6 | 0 | 1[#] | | | 1[#] | | | |
| MODERN | Hawaii | Kokee | 11 | 10 | 0 | | 10 | | | | |
| | Marquesas | - | 6 | 5 | 0 | 3 | 2 | | | | |
| | Vanuatu | Efate | | 14 | | 11 | 2 | 1 | | | Dancause *et al.* (10) |
| | | Tanna | | 3 | | 3 | | | | | |
| | | Aneityum | | 19 | | 19 | | | | | |
| | | Ambae | | 7 | | 7 | | | | | |
| | Guam | - | | 5 | | 3 | 2 | | | | |
| | Santa Cruz | Anuta Island | | 7 | | 7 | | | | | This study |
| | | Tikopia | | 7 | | 5 | 2 | | | | |
| | | Nendo Island | | 5 | | 5 | | | | | |
| | | Vanikoro Island | | 5 | | 5 | | | | | |
| | | Temotu Neo | | 2 | | 2 | | | | | |
| | | Utupua Island | | 2 | | 2 | | | | | |
| | Solomon Islands | Nggela Island | | 9 | | 7 | 2 | | | | |
| | | Makira-Ulawa | | 6 | | 6 | | | | | |
| | | Rendova Island | | 6 | | 5 | | | 1 | | |
| | | Russell Island | | 5 | | 5 | | | | | |
| | | Treasury Island | | 5 | | 5 | | | | | |
| | Papua New Guinea | Karkar Island | | 6 | | 5 | 1 | | | | |
| | | Witu Island | | 4 | | 4 | | | | | |
| | | Watom Island | | 3 | | 3 | | | | | |
| | Indonesia | Baik Island | | 4 | | 4 | | | | | |
| | | Mulia | | 2 | | 2 | | | | | |
| | | Enarotali | | 1 | | | 1 | | | | |
| | | Nabire | | 1 | | 1 | | | | | |
| | | Ternate Island | | 1 | | 1 | | | | | |
| | | Wamena | | 1 | | 1 | | | | | |
| | Philippines | Camiguin | | 8 | | 3 | 4 | | | 1 | |
| | | Pintuyan Town | | 4 | | 3 | 1 | | | | |
| | | Jagna Province | | 3 | | 2 | 1 | | | | |
| | | Cebu City | | 2 | | | 2 | | | | |
| | | Balicasag Island | | 1 | | | 1 | | | | |
| | | Palawan | | 1 | | 1 | | | | | |
| | | Pamalican Island | | 1 | | 1 | | | | | |
| | | Panglao Island | | 1 | | 1 | | | | | |
| | | Merlia Farm | | 1 | | | | 1 | | | |
| | | Unknown | | 1 | | 1 | | | | | |
| | Vietnam | Ho Chi Minh | | 2 | | | | | 2 | | |

* Haplotype designation from Liu *et al.* (13)

[#] Results not replicable and identified as contamination after Shrimp DNase treatment.

Table S2. Data used in Bayesian Serial Simcoal (BayeSSC) analysis for investigating the origins of South American chickens. This dataset uses ancient samples from all haplogroups i.e. it includes samples from all haplogroups (this paper, plus those amplified without Shrimp DNase: 1, 25, 26, 41).

| Age of sample (years BP) | Sequences included | Region | Temporal scale | Site | Reference |
|---|---|---|---|---|---|
| 0 | 795 sequences | MSEA | Modern | China, Laos, Myanmar, Vietnam, Thailand, Malaysia | (9, 13, 42) |
| 2700 | THABNW003 | MSEA | Ancient | Ban Non Wat site, Thailand | (25) |
| 0 | 330 sequences | ISEA | Modern | Philippines, Japan, Indonesia, New Guinea | (13, 16) |
| 0 | 58 sequences | Europe | Modern | Europe | (13) |
| 980 | ESPALB002 | Europe | Ancient | Albarracin (Teruel) site, Spain | (25) |
| 350 | ESPLCT001 | Europe | Ancient | La Cartuja (Seville) site, Spain | (25) |
| 0 | 39 sequences | South America | Modern | Chile, | (11) |
| 350 | PRUTOR001 | South America | Ancient | Torata Alta site, Peru | (25) |
| 657 | CHLARA001 | South America | Ancient | El Arenal-1 site, Chile | (1) |
| 536 | CHLARA004 | South America | Ancient | El Arenal-1 site, Chile | (41) |
| 350 | PRULOC001 | South America | Ancient | Locumbilla Winery site, Peru | (25) |
| 540 | CHLARA003 | South America | Ancient | El Arenal-1 site, Chile | (41) |
| 0 | 122 sequences | Pacific | Modern | Guam, Solomon Islands, Santa Cruz, Vanuatu, Marquesas, Hawaii | this paper |
| 660 | ACAD9057, ACAD9067 | Pacific | Ancient | Anakena site, Rapa Nui | this paper |
| 680 | ACAD9066, ACAD9068, ACAD9070, ACAD9072, ACAD9073, ACAD9074 | Pacific | Ancient | Anakena site, Rapa Nui | this paper |
| 700 | PAQANA010, ACAD9071 | Pacific | Ancient | Anakena site, Rapa Nui | (this paper; 1) |
| 700 | ACAD9069, ACAD9065 | Pacific | Ancient | Anakena site, Rapa Nui | This paper |
| 670 | PAQANA006 | Pacific | Ancient | Anakena site, Rapa Nui | (1) |
| 600 | ACAD9075, ACAD9076 | Pacific | Ancient | Anakena site, Rapa Nui | this paper |
| 534 | ACAD8136, ACAD8668, ACAD8670, ACAD8671, ACAD8672, ACAD8674, ACAD8675 | Pacific | Ancient | Makauwahi cave site, Hawaii | this paper |
| 1285 | ACAD3895, ACAD3896 | Pacific | Ancient | Anatoloa site, Niue | this paper |
| 2000 | Tonga HB, Tonga TD | Pacific | Ancient | Mele Havea site; Tongoleleka site, Tonga | (1) |
| 1000 | ASMFTF001 | Pacific | Ancient | Fatu-ma-Futi site, American Samoa | (1) |
| 810 | FSMFSP001 | Pacific | Ancient | FSPO-4 site, Fais | (25) |
| 2775 | SLB33001 | Pacific | Ancient | SE-SZ-33 site, Santa Cruz | (25) |
| 228 | PAQHAN001 | Pacific | Ancient | Hangahahave site, Rapa Nui | (1) |
| 1550 | FSMFSP002, FSMFSP003 | Pacific | Ancient | FSPO-8 site; FSPO-4 site, Fais | (25) |
| 2605 | VUTTEO003 | Pacific | Ancient | Teouma site, Vanuatu | (26) |
| 910 | HWIKUA001 | Pacific | Ancient | Kualoa, O'ahu, Hawaii | (1) |
| 1590 | NIUPKI009 | Pacific | Ancient | Paluki site, Niue | (1) |
| 2974 | VUTTEO006 | Pacific | Ancient | Teouma site, Vanuatu | (26) |

Table S3. Data used in Bayesian Serial Simcoal (BayeSSC) analysis for investigating the origins of South American chickens. This dataset uses ancient samples from only haplogroup D (this paper; 1, 25).

| Age of sample (years BP) | Sequences included | Region | Temporal scale | Site | Reference |
|---|---|---|---|---|---|
| 0 | 795 sequences | MSEA | Modern | China, Laos, Myanmar, Vietnam, Thailand, Malaysia | (9, 13, 42) |
| 0 | 330 sequences | ISEA | Modern | Philippines, Japan, Indonesia, New Guinea | (13, 16) |
| 0 | 58 sequences | Europe | Modern | Europe | (13) |
| 0 | 39 sequences | South America | Modern | Chile, | (11) |
| 350 | PRUTOR001 | South America | Ancient | Torata Alta site, Peru | (25) |
| 0 | 122 sequences | Pacific | Modern | Guam, Solomon Islands, Santa Cruz, Vanuatu, Marquesas, Hawaii | this paper |
| 660 | ACAD9057, ACAD9067 | Pacific | Ancient | Anakena site, Rapa Nui | this paper |
| 680 | ACAD9066, ACAD9068, ACAD9070, ACAD9072, ACAD9073, ACAD9074 | Pacific | Ancient | Anakena site, Rapa Nui | this paper |
| 700 | PAQANA010, ACAD9071 | Pacific | Ancient | Anakena site, Rapa Nui | (this paper; 1) |
| 700 | ACAD9069, ACAD9065 | Pacific | Ancient | Anakena site, Rapa Nui | This paper |
| 670 | PAQANA006 | Pacific | Ancient | Anakena site, Rapa Nui | (1) |
| 600 | ACAD9075, ACAD9076 | Pacific | Ancient | Anakena site, Rapa Nui | this paper |
| 534 | ACAD8136, ACAD8668, ACAD8670, ACAD8671, ACAD8672, ACAD8674, ACAD8675 | Pacific | Ancient | Makauwahi cave site, Hawaii | this paper |
| 1285 | ACAD3895, ACAD3896 | Pacific | Ancient | Anatoloa site, Niue | this paper |
| 810 | FSMFSP001 | Pacific | Ancient | FSPO-4 site, Fais | (25) |
| 228 | PAQHAN001 | Pacific | Ancient | Hangahahave site, Rapa Nui | (1) |

Table S4. Migration matrix used in the BayeSSC investigation of the origins of South American chickens (ancient samples from all haplogroups). The matrix represents the ratio of the lineages in each row that will migrate to each column backwards through time (i.e. reconstructed via the coalescence).

| | MSEA | ISEA | Europe | South America | Pacific |
|---|---|---|---|---|---|
| MSEA | 0 | 0 | 0.00001 | 0 | 0 |
| ISEA | 0.00001 | 0 | 0.00001 | 0 | 0.00001 |
| Europe | 0.00001 | 0 | 0 | 0.0000001 | 0 |
| South America | 0 | 0 | 0.0001 | 0 | 0 |
| Pacific | 0 | 0.00001 | 0.0001 | 0.0000001 | 0 |

Table S5. Migration matrix used in the BayeSSC investigation of the origins of South American chickens for models via Europe (D haplogroup ancient samples only). The matrix represents the ratio of the lineages in each row that will migrate to each column backwards through time (i.e. reconstructed via the coalescence).

| | MSEA | ISEA | Europe | SthAm | Pacific |
|---|---|---|---|---|---|
| MSEA | 0 | 0 | 0.00001 | 0 | 0 |
| ISEA | 0.00001 | 0 | 0.00001 | 0 | 0.00001 |
| Europe | 0.00001 | 0 | 0 | 0.0000001 | 0 |
| SthAm | 0 | 0 | 0.0001 | 0 | 0 |
| Pacific | 0 | 0.00001 | 0.0001 | 0.0000001 | 0 |

Table S6. $F_{ST}$ population differentiation statistics for all modern haplogroup D populations

| n | | Japan | Indonesia | Philippines | PNG | Guam | Santa Cruz | Solomon Island | Vanuatu | Marquesas | Vietnam | China | Thailand | Myanmar |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 33 | Japan | 0 | | | | | | | | | | | | |
| 19 | Indonesia | 0.19** | 0 | | | | | | | | | | | |
| 12 | Philippines | 0.15* | 0.03 | 0 | | | | | | | | | | |
| 12 | PNG | 0.35** | 0.50** | 0.26** | 0 | | | | | | | | | |
| 3 | Guam | 0.09 | 0.10 | -0.05 | 0.59** | 0 | | | | | | | | |
| 26 | SantaCruz | 0.28** | 0.22** | 0.13** | 0.22** | 0.14 | 0 | | | | | | | |
| 28 | SolomonIs | 0.29** | 0.21** | 0.13** | 0.20** | 0.15 | -0.01 | 0 | | | | | | |
| 40 | Vanuatu | 0.31** | 0.34** | 0.24** | 0.34** | 0.27** | 0.08** | 0.11** | 0 | | | | | |
| 3 | Marquesas | 0.23** | 0.51** | 0.15* | 1.00** | 0.25 | 0.02 | 0.10 | -0.08 | 0 | | | | |
| 2 | Vietnam | -0.08 | -0.20 | -0.24 | 0.94 | -0.29 | 0.14 | 0.16 | 0.24 | 0.77 | 0 | | | |
| 23 | China | 0.13** | 0.25** | 0.16** | 0.44** | 0.10 | 0.36** | 0.37** | 0.38 | 0.30 | -0.04 | 0 | | |
| 2 | Thailand | 0.05 | 0.07 | -0.09 | 0.94** | -0.13 | 0.31 | 0.34 | 0.44 | 0.86 | 0.00 | 0.09 | 0 | |
| 2 | Myanmar | 0.36** | 0.77** | 0.45** | 1.00* | 0.45 | 0.69** | 0.71** | 0.67 | 1.00 | 0.86 | 0.13 | 0.89 | 0 |

* p-value < 0.05; ** p-value < 0.01.

Table S7. Data used in Bayesian Serial Simcoal (BayeSSC) analysis for investigating the prehistoric colonization history of chickens in Micronesia. This dataset uses ancient samples from only haplogroup D (it includes samples from both this paper and Storey *et al.* (1, 25)).

| Age of sample (years BP) | 15 sample groups | Region | Temporal scale | Site | Reference |
|---|---|---|---|---|---|
| 0 | 15 samples | Philippines | Modern | Philippines | (this paper; 9, 13, 42) |
| 0 | 15 samples | NG | Modern | PNG | this paper |
| 0 | 5 samples | Micronesia | Modern | Micronesia | (10) |
| 0 | 32 samples | Solomons | Modern | Solomon Islands | this paper |
| 0 | 71 samples | Western Polynesia | Modern | Santa Cruz & Vanuatu | this paper |
| 0 | 14 samples | Eastern Polynesia | Modern | Hawaii & Marquesas | this paper |
| 810 | FSMFSP001 | Micronesia | Ancient | FSPO-4 site, Fais | (25) |
| 1285 | 2 samples | Central Polynesia | Ancient | Niue | this paper |
| 660 | ACAD9057, ACAD9067 | Eastern Polynesia | Ancient | Anakena site, Rapa Nui | this paper |
| 680 | ACAD9066, ACAD9068, ACAD9070, ACAD9072, ACAD9073, ACAD9074 | Eastern Polynesia | Ancient | Anakena site, Rapa Nui | this paper |
| 700 | PAQANA010, ACAD9071 | Eastern Polynesia | Ancient | Anakena site, Rapa Nui | (this paper; 1) |
| 700 | ACAD9069, ACAD9065 | Eastern Polynesia | Ancient | Anakena site, Rapa Nui | this paper |
| 670 | PAQANA006 | Eastern Polynesia | Ancient | Anakena site, Rapa Nui | (1) |
| 600 | ACAD9075, ACAD9076 | Eastern Polynesia | Ancient | Anakena site, Rapa Nui | this paper |
| 534 | ACAD8136, ACAD8668, ACAD8670, ACAD8671, ACAD8672, ACAD8674, ACAD8675 | Eastern Polynesia | Ancient | Makauwahi cave site, Hawaii | this paper |

Table S8. Migration matrix used in the BayeSSC investigating the prehistoric colonization history of chickens in Micronesia (D haplogroup ancient samples only). The matrix represents the ratio of the lineages in each row that will migrate to each column backwards through time (i.e. reconstructed via the coalescence).

| | Philippines | New Guinea | Micronesia | Solomon Islands | Western Polynesia | Eastern Polynesia |
|---|---|---|---|---|---|---|
| Philippines | 0 | 0 | 0.00001 | 0 | 0 | 0 |
| New Guinea | 0.00001 | 0 | 0.00001 | 0.00001 | 0 | 0 |
| Micronesia | 0.00001 | 0.00001 | 0 | 0.00001 | 0 | 0 |
| Solomon Islands | 0 | 0.0001 | 0.00001 | 0 | 0.00001 | 0 |
| Western Polynesia | 0 | 0.000001 | 0 | 0.00001 | 0 | 0.00001 |
| Eastern Polynesia | 0 | 0 | 0 | 0 | 0.00001 | 0 |

Table S9. Population genetic summary statistics for haplogroup D in the Asia-Pacific region.

| | n | #Hap | Hdiv | nDiv (%) | Ts | Tv | Tajima's D | Fu's FS |
|---|---|---|---|---|---|---|---|---|
| Vietnam | 2 | 2 | 1.00 | 0.50 | 1 | 1 | 0.00 | 0.00 |
| Thailand | 2 | 2 | 1.00 | 0.50 | 1 | 0 | 0.00 | 0.00 |
| Guam | 3 | 3 | 1.00 | 1.99 | 6 | 0 | 0.00 | 0.13 |
| Philippines | 12 | 6 | 0.89 | 1.65 | 12 | 0 | -0.69 | -1.09 |
| Japan | 33 | 8 | 0.81 | 2.08 | 11 | 1 | 1.33 | 1.50 |
| Santa Cruz | 26 | 6 | 0.81 | 0.83 | 6 | 0 | 0.19 | -0.35 |
| China | 23 | 7 | 0.78 | 1.83 | 1 | 1 | 1.68 | 0.97 |
| Vanuatu | 40 | 8 | 0.76 | 0.81 | 9 | 1 | -0.92 | -1.51 |
| Indonesia | 19 | 6 | 0.74 | 0.53 | 4 | 0 | -0.21 | -2.16 |
| Solomon Islands | 28 | 4 | 0.71 | 0.76 | 4 | 0 | 1.25 | 1.45 |
| Myanmar | 2 | 1 | 0.00 | 0.00 | 0 | 0 | 0 | - |
| Marquesas | 3 | 1 | 0.00 | 0.00 | 0 | 0 | 0 | - |
| PNG | 12 | 1 | 0.00 | 0.00 | 0 | 0 | 0 | - |

n – number of samples; #Hap – number of haplotypes; Hdiv – haplotype diversity; nDiv – nucleotide diversity; Ts – transitions; Tv – transversions.

Table S10. Primer sequences

| Primer name | Primer sequence (5’ to 3’) | Reference |
|---|---|---|
| GG144F | ACCCATTATATGTATACGGGCATTAA | (1) |
| GG387R | CGAGCATAACCAAATGGGTTAGA | (1) |
| GG316F | AACAAGTCACCTAACTATGAATGGTTAC | (1) |
| GG586R | AGTTATGCATGGGATGTGCCTGACCGA | (1) |
| A1780F | CAGCTCCAAACCACTACCAAG | This paper |
| A1781R | AGGTGACTTGTTGGGGGAAG | This paper |
| A1958F | TCTAACTCATTTGGTTATGCTCG | This paper |
| A1959R | AGTTATGTATGGGATGTGCCTGACCGA | This paper |

***Supplementary Figures***

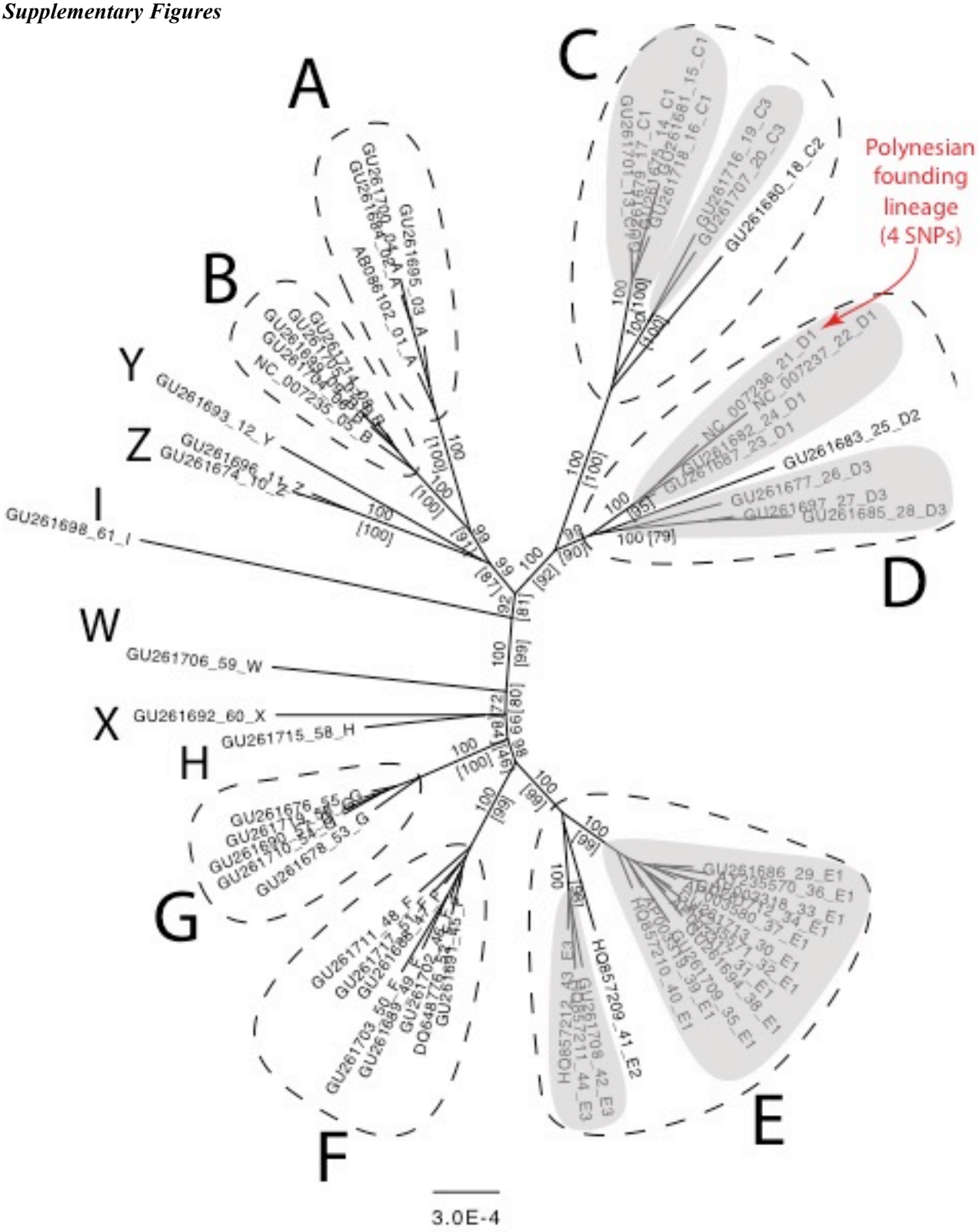


Fig. S1. Bayesian phylogenetic tree based on the whole mitochondrial genome dataset of 61 Miao *et al.* (43) excluding the control region. Numbers above branches are Bayesian posterior probabilities, with Maximum Likelihood bootstrap values shown below branches in square brackets.

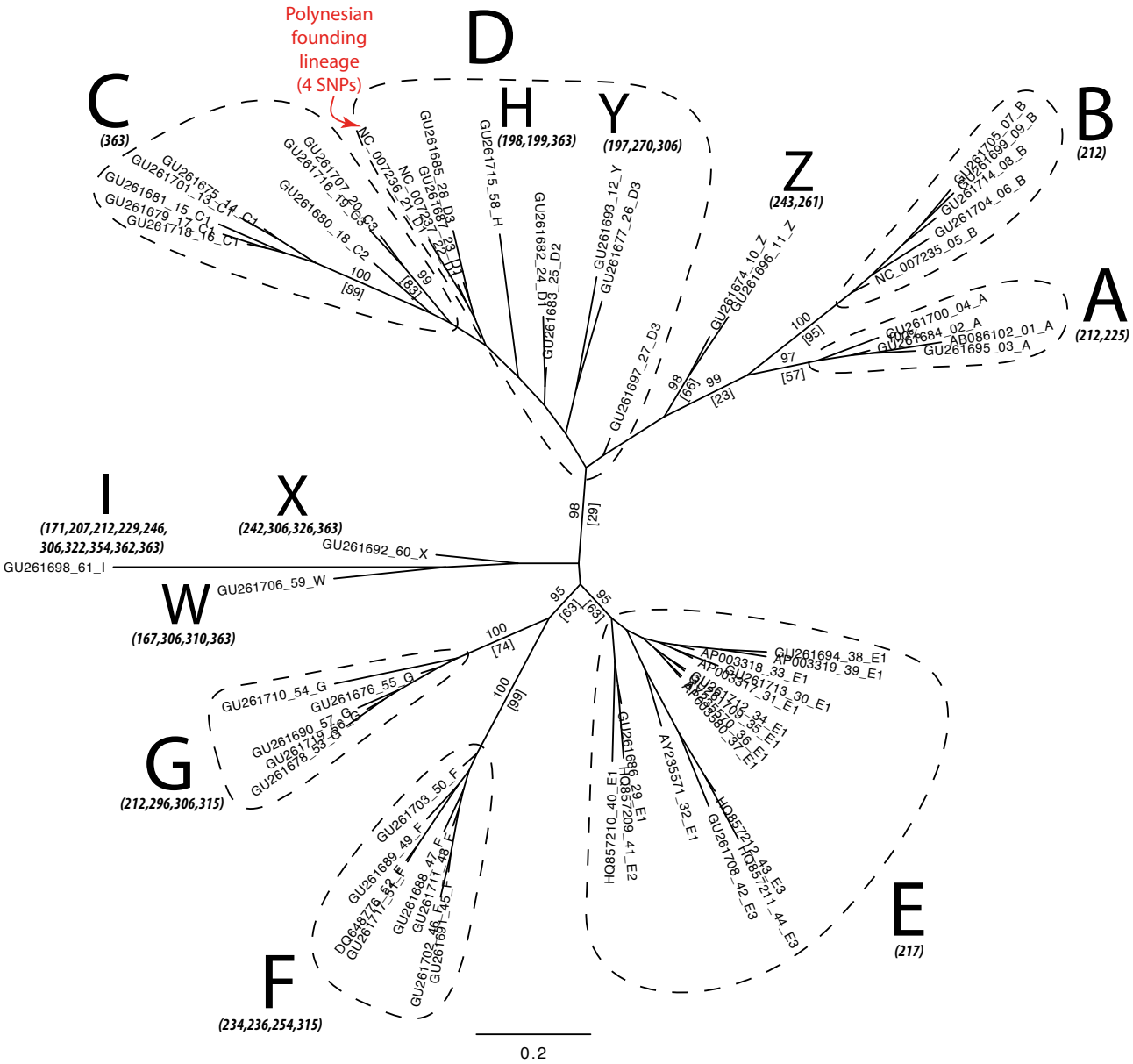


Fig. S2. Bayesian phylogenetic tree based on 201bp fragment of the control region from 61 WMG sequences of Miao *et al.* (43). Numbers above branches are Bayesian posterior probabilities, with Maximum Likelihood bootstrap values shown below branches in square brackets. The WMG sequence that contains all 4 ancestral Polynesian SNPs is highlighted in red. Below each haplogroup label are the nucleotide positions that differentiate the haplogroup from Haplogroup D (*i.e.* Haplogroup D defining SNPs).

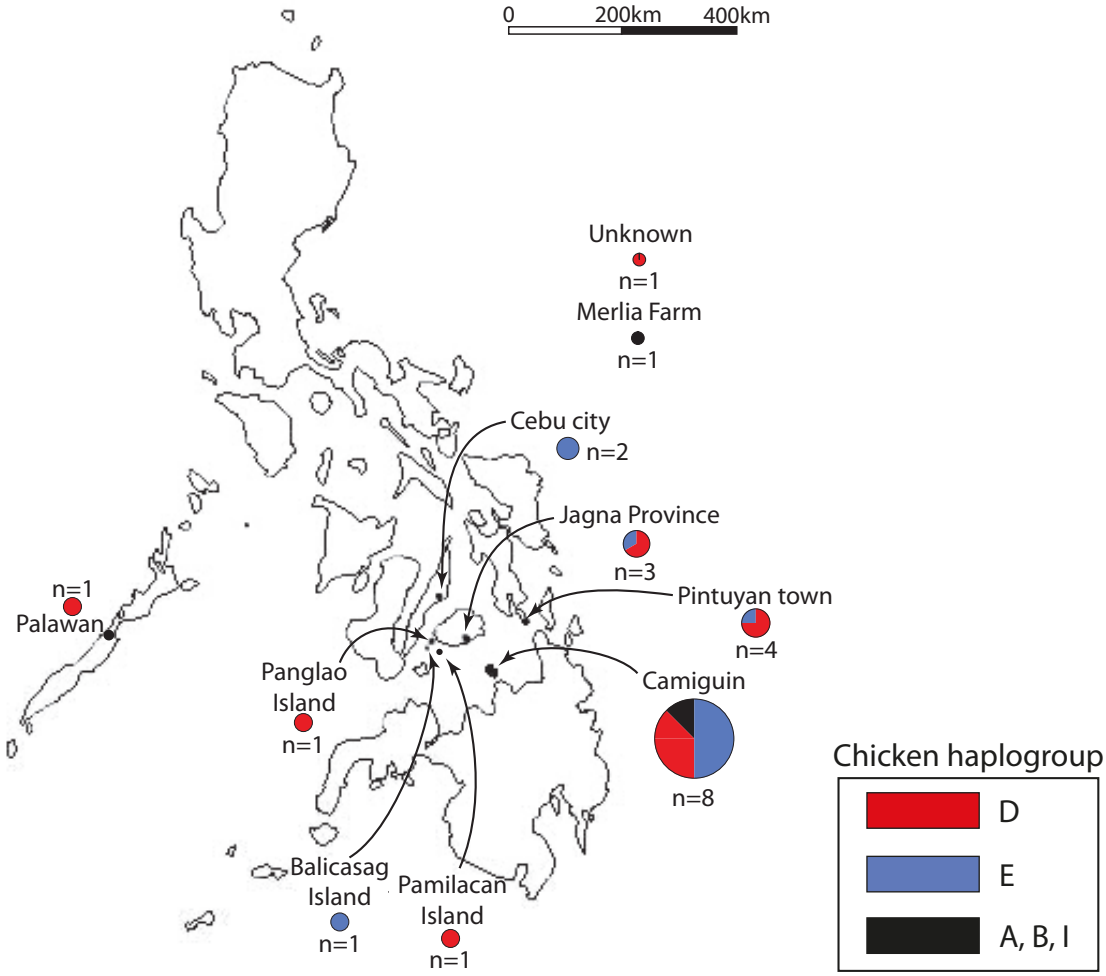


Fig. S3. Map showing haplogroups of chicken samples from the Philippines (n=23).


Fig. S4. Map showing haplogroups of chicken samples from the Indonesia (n=10).

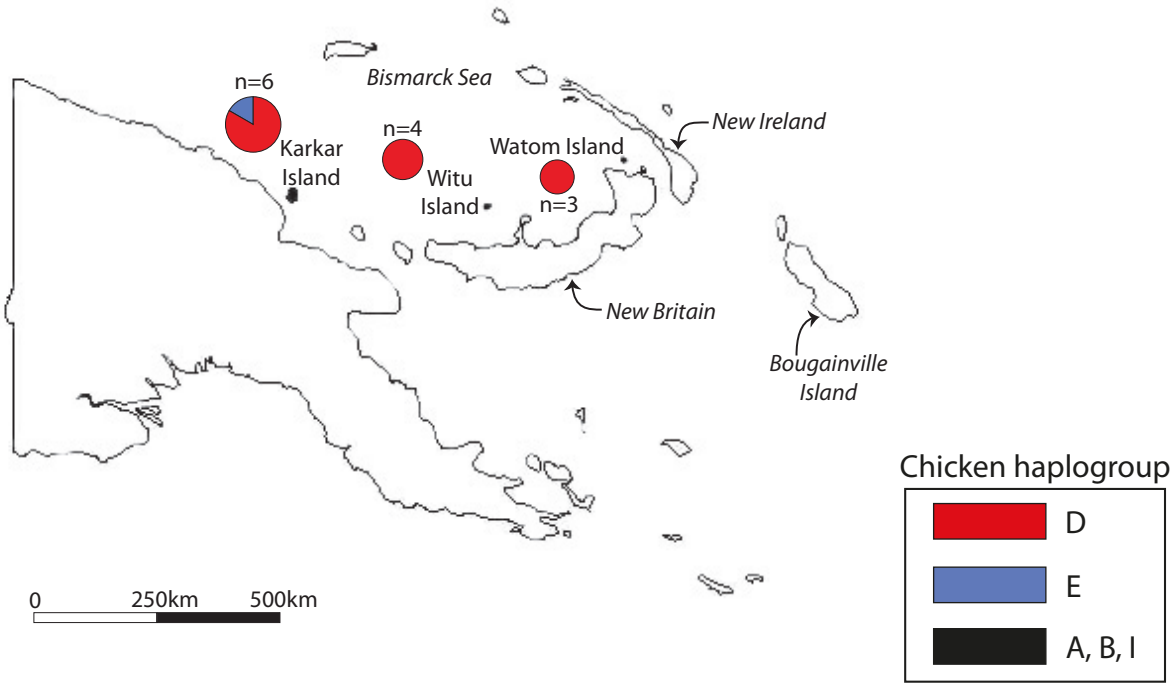


Fig. S5. Map showing haplogroups of chicken samples from the Papua New Guinea (n=13).

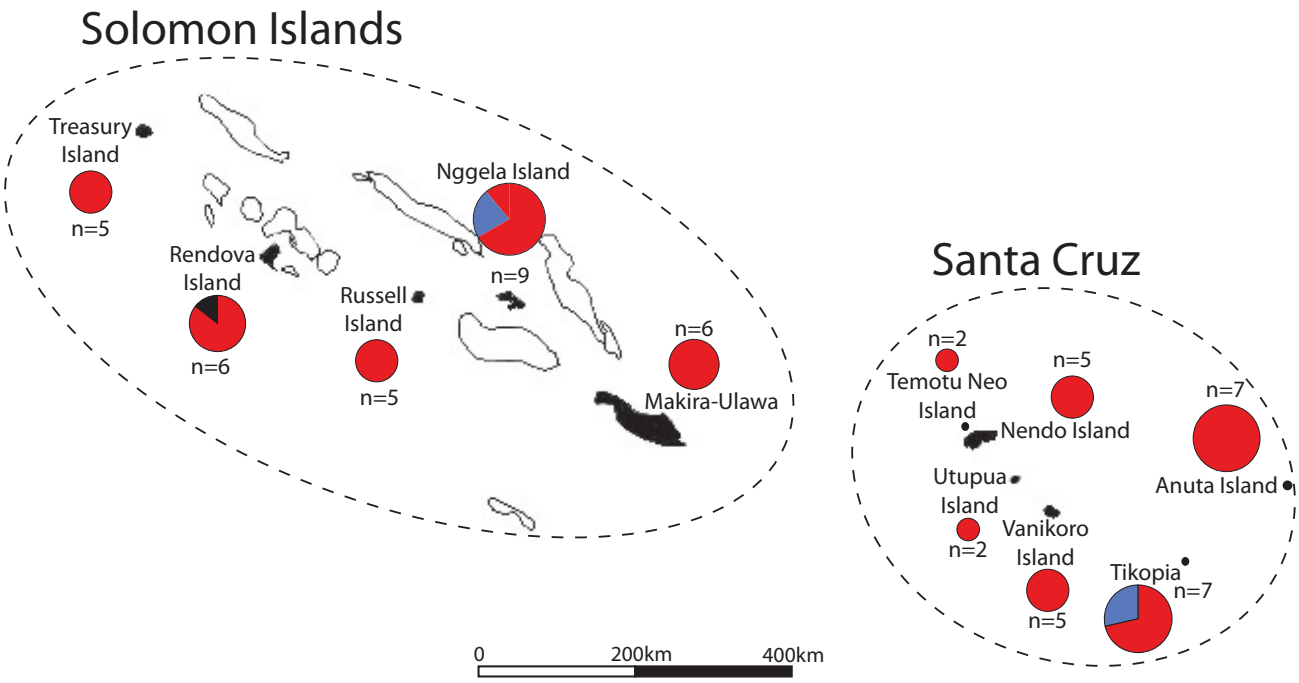


Fig. S6. Map showing haplogroups of chicken samples from the Solomon Islands (n=31) and Santa Cruz (n=28).

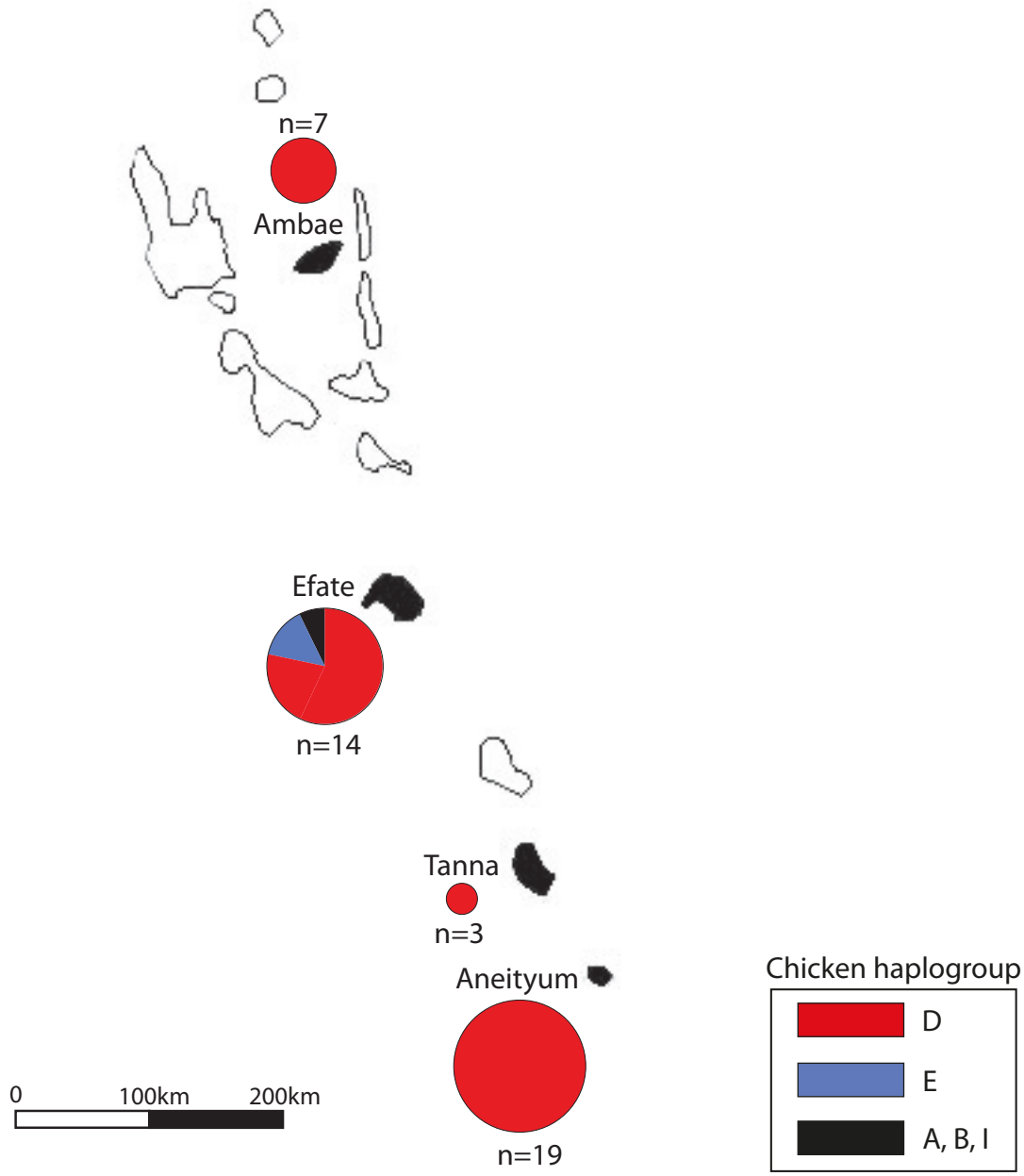


Fig. S7. Map showing haplogroups of chicken samples from the Vanuatu (n=43), Dancause *et al.* (2010).

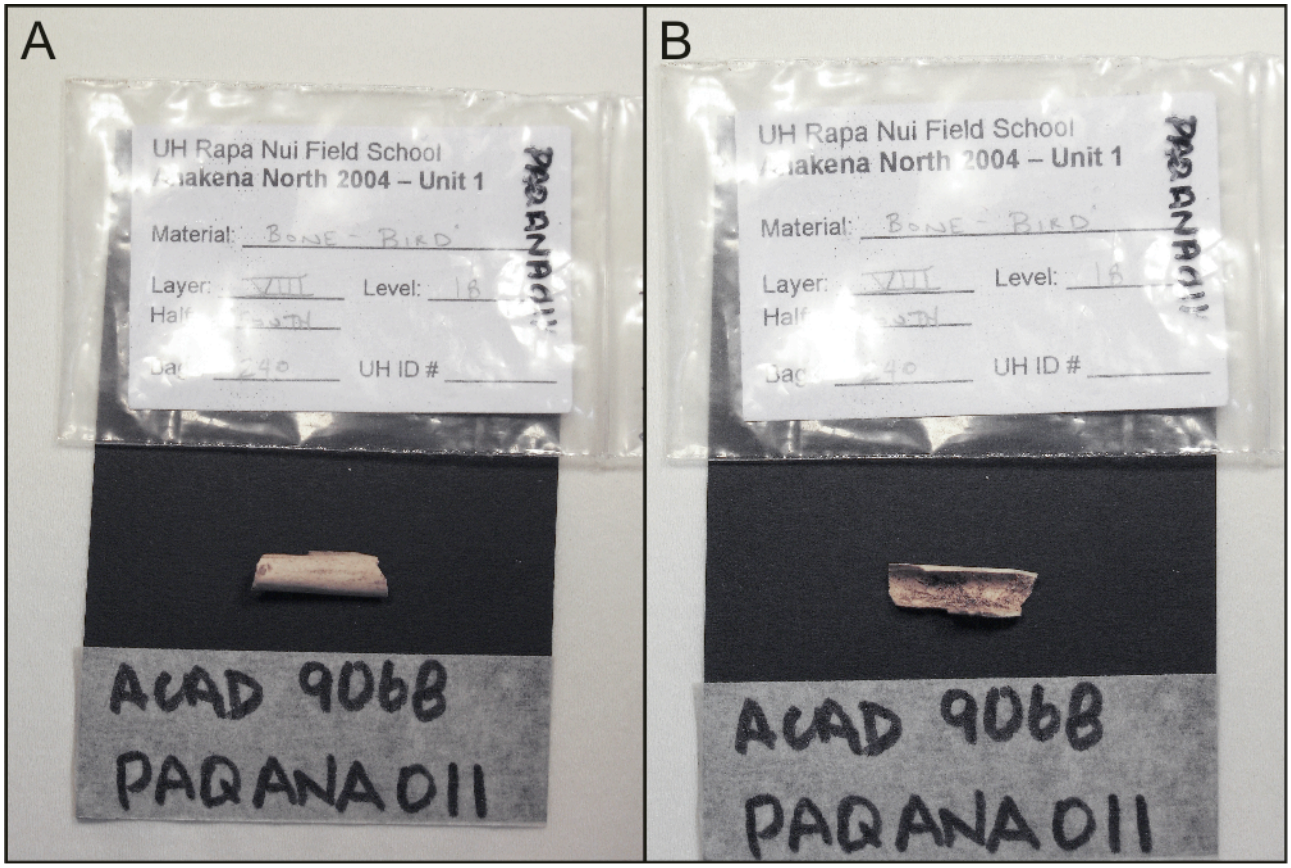


Fig. S8. Photos showing sample PAQANA011, the sample from Rapa Nui that Storey *et al.* (1) found to be Haplogroup E but when re-analyzed for this paper at ACAD and replicated at Durham was actually found to be Haplogroup D. Photo A shows the exterior of the bone and photo B shows the interior of the bone upon arrival at the ACAD lab (*i.e.* prior to subsampling for analysis and replication).

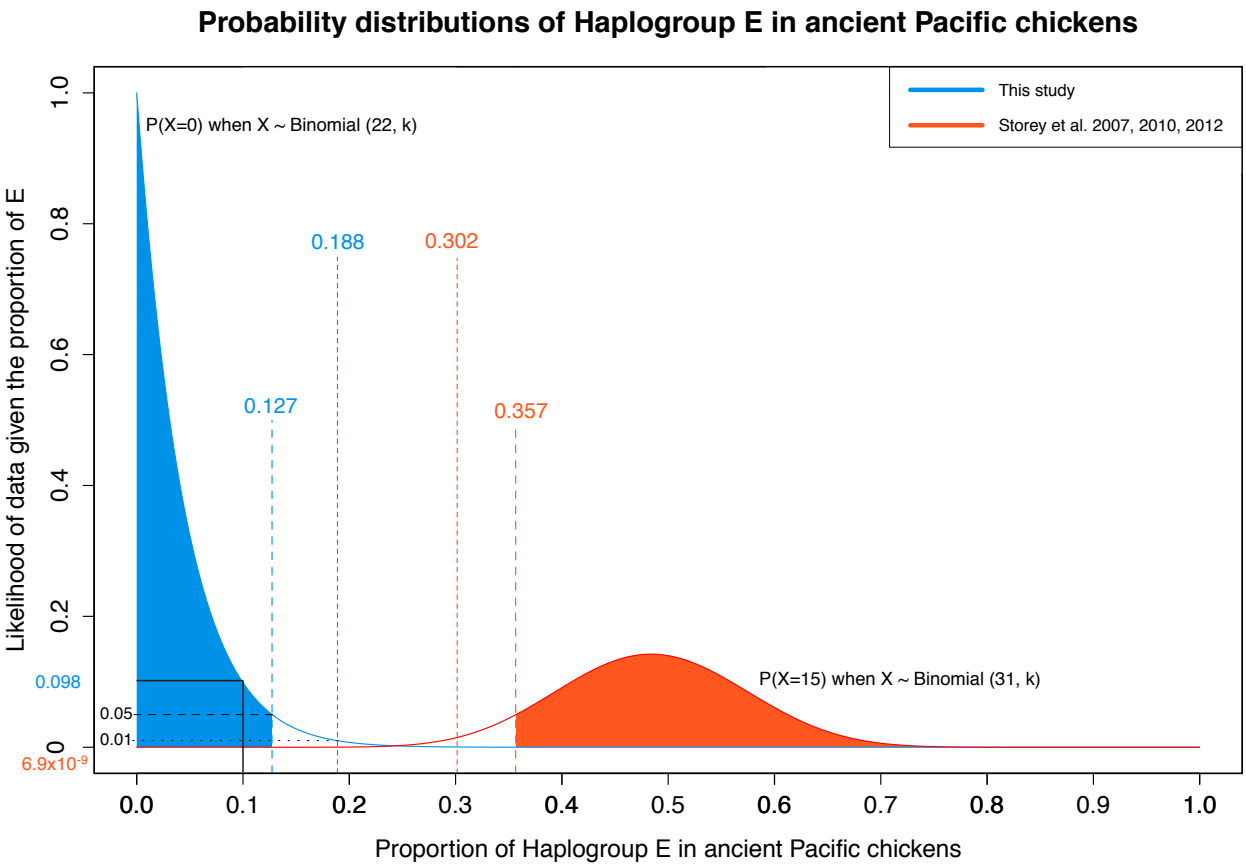


Fig. S9. Binomial probability distribution showing the probability of detecting the observed number of haplogroup E sequences (for a range of hypothetical frequencies of E in the ancient Pacific chicken population). For example, if we assume that haplogroup E is actually present at a frequency of 0.1 (i.e. 10%) in the total ancient population, then the probability of detecting 0/22 haplogroup E sequences is 0.098 (this study), but the probability of Storey *et al.* (1, 25, 26) detecting 15/31 haplogroup E sequences is $6.9 \times 10^{-9}$.

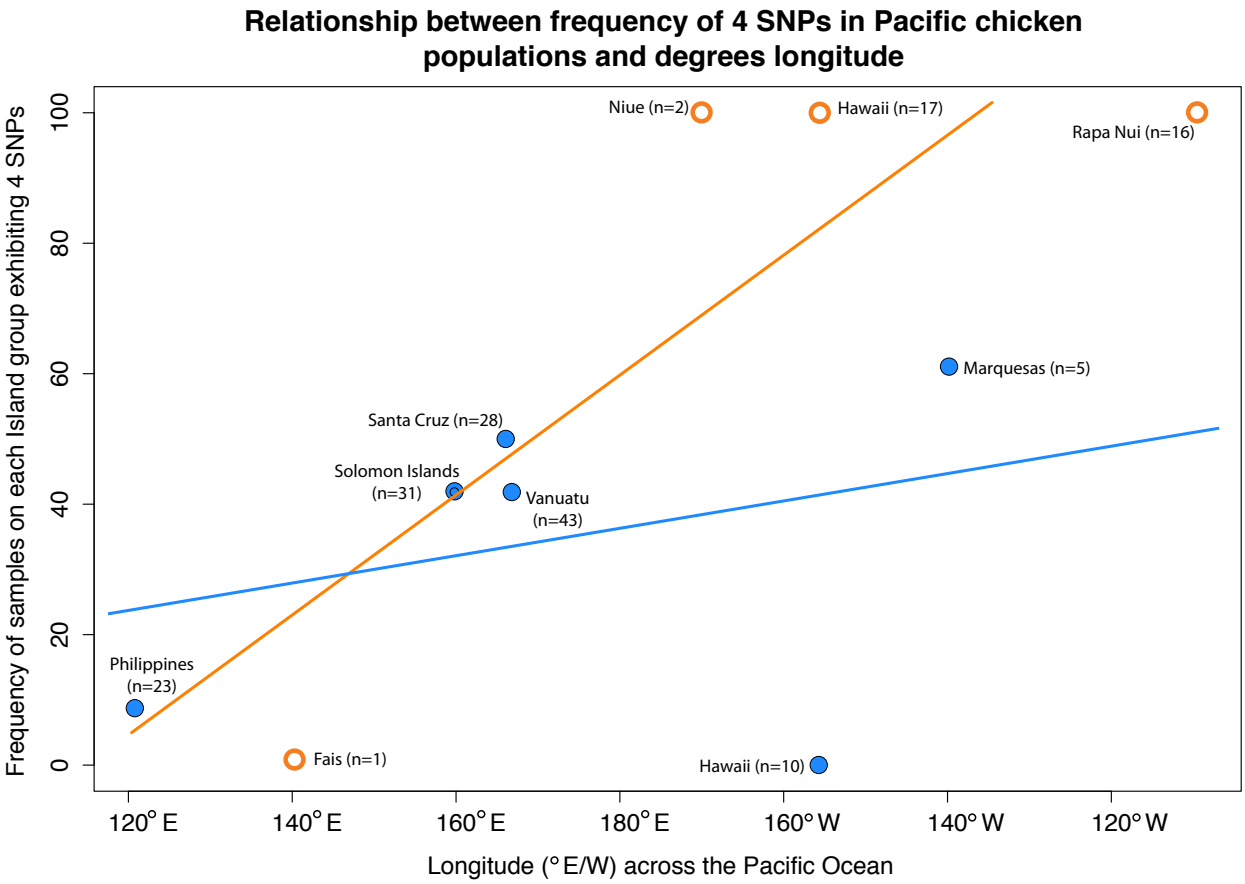


Fig. S10. Linear regression plot showing the positive correlation between frequency of haplotypes with all four diagnostic SNPs and longitude. The higher frequency of these 4 SNPs in the east (right hand side of the graph) is apparent with both modern (blue solid) and ancient (orange outline) samples. With both modern and ancient datasets combined, the correlation is relatively low but the linear relationship is statistically significant ($R^2$=0.4201 and ANOVA P-value = 0.04 for modern and ancient combined). Each island group is labeled with sample numbers in brackets, but note that no ancient haplogroup E samples are included due to suspected issues with authenticity.

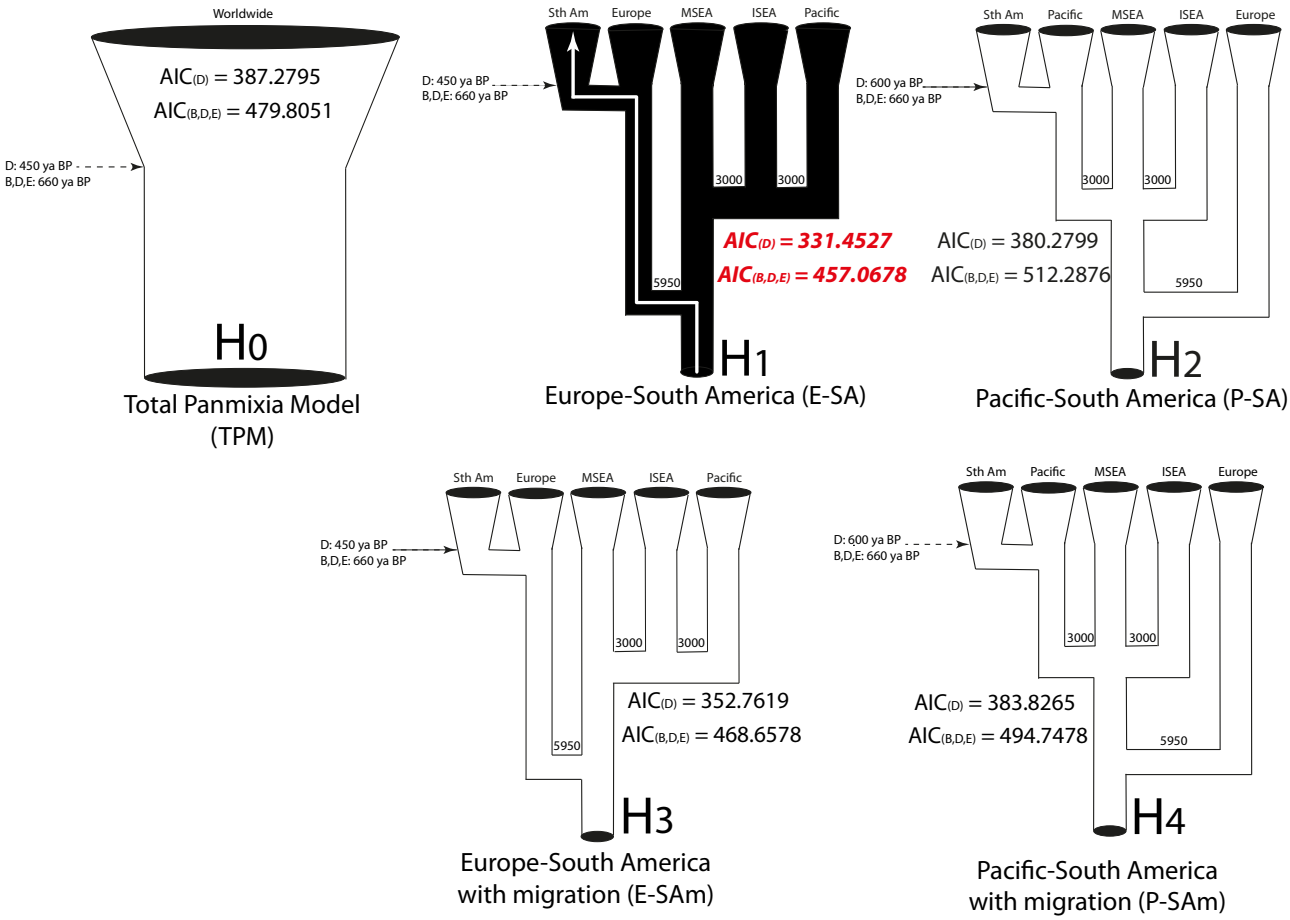


Fig. S11. Serial Coalescent Simulations and Approximate Bayesian Computation models with respect to reconstructing the origin of South American chicken populations. The null hypothesis was modeled as the Total Panmixia Model, TPM (H0). Each population (South America, Europe, MSEA, ISEA, Pacific) was sampled at random from a panmictic population. Eight alternate scenarios were also tested: route from Europe-South America with or without migration, H3 or H1 respectively, with each model having two variations, based on different datasets (only D haplogroup ancient samples or B, D and E haplogroup ancient samples), and a route from Pacific-South America with or without migration, H4 or H2 respectively, with each model having two variations, based on different datasets (only D haplogroup ancient samples or B, D and E haplogroup ancient samples).

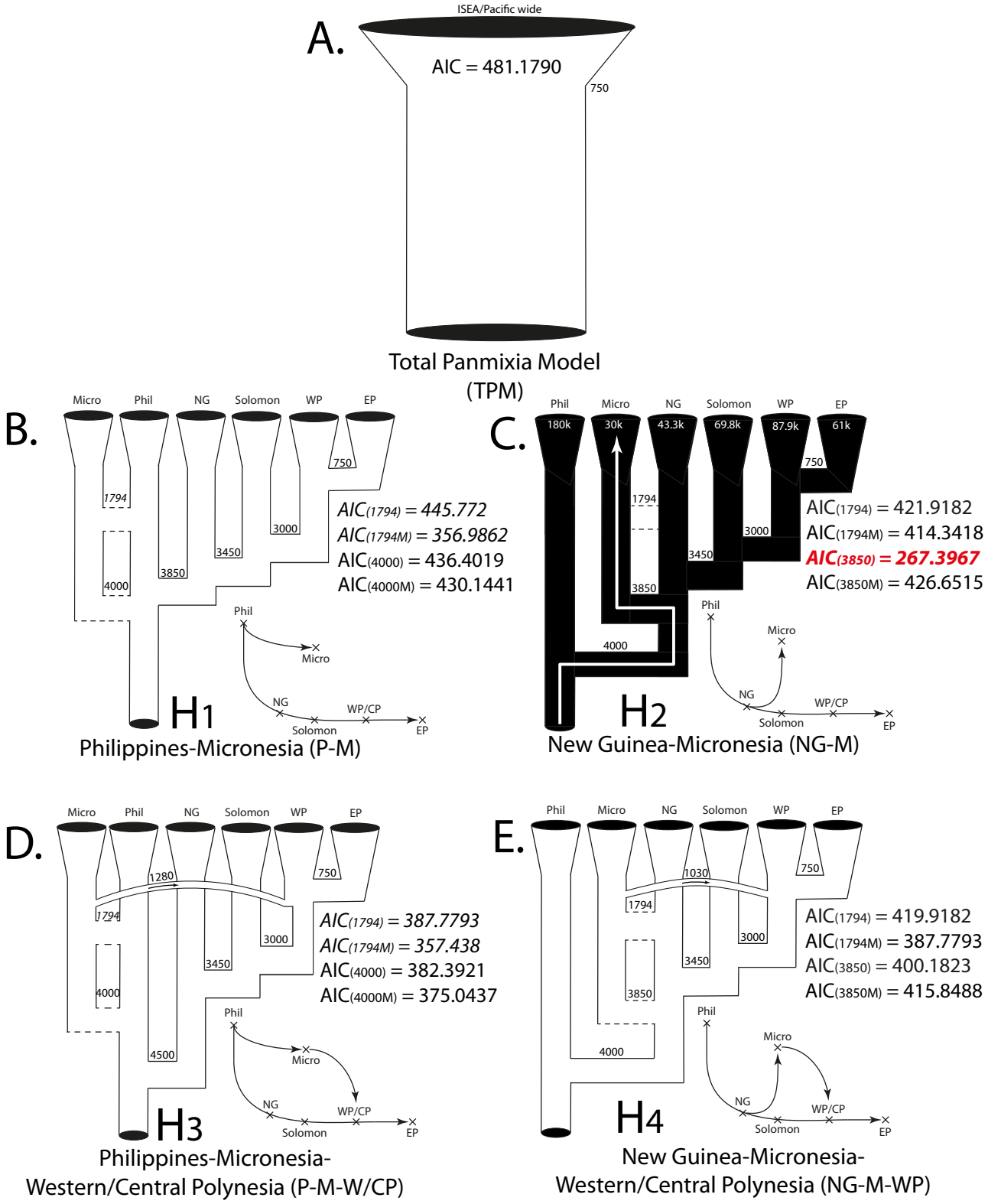


Fig. S12. Serial Coalescent Simulations and Approximate Bayesian Computation models with respect to how Micronesia fits into the wider Pacific story. (A) Total Panmixia Model, TPM. Each population (the Philippines, New Guinea, Micronesia, the Solomon Islands, Western/Central Polynesia, and Eastern Polynesia) was sampled at random from a panmictic population. (B) Philippines-Micronesia model, P-M. This model has four variations, based on two temporal versions for the migration from the Philippines to Micronesia (1794 yrs BP or 4000 yrs BP; dotted lines) and two levels of migration since 750 yrs BP (no migration or a migration matrix; see Table S8). (C) New Guinea-Micronesia, NG-M. This model has four variations, based on two temporal versions for the migration from New Guinea to Micronesia (1794 yrs BP or 3850 yrs BP) and two levels of migration since 750 yrs BP (no migration or a migration matrix; see Table S8). (D) Philippines-Micronesia-Western/Central Polynesia, P-M-W/CP. This model includes a percentage of migration from Micronesia to Western Central Polynesia based on a prior uniform distribution ranging from 750-1794 yrs BP. Again this model has four variations, based on two temporal versions for the

migration from the Philippines to Micronesia (1794 yrs BP or 4000 yrs BP) and two levels of migration since 750 yrs BP (no migration or a migration matrix; see Table S8). (E) New Guinea-Micronesia-Western/Central Polynesia, NG-M-W/CP. This model includes a percentage of migration from Micronesia to Western/Central Polynesia based on a prior uniform distribution ranging from 750-1794 yrs BP. Again this model has four variations, based on two temporal versions for the migration from the Philippines to Micronesia (1794 yrs BP or 4000 yrs BP) and two levels of migration since 750 yrs BP (no migration or a migration matrix; see Table S8).

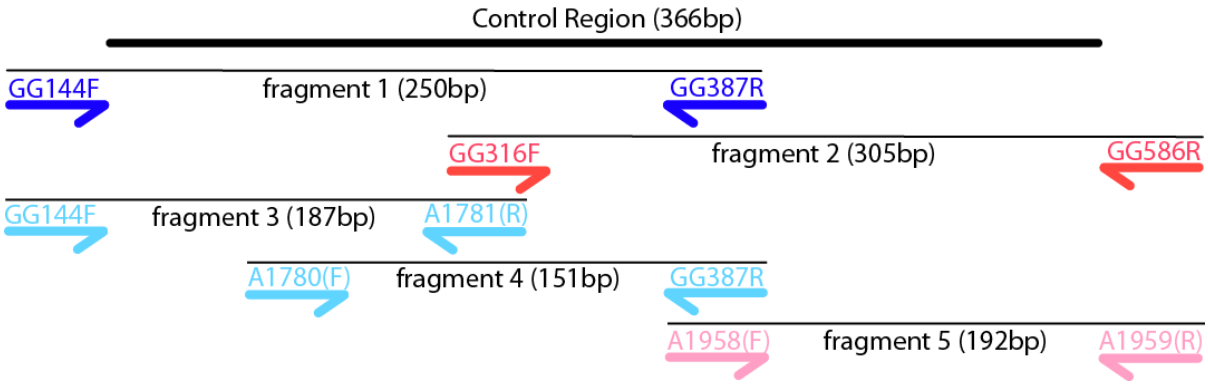


Fig. S13. Details of primer arrangement showing the 366 bp target region, prior to trimming sequences to the length common across both ancient and modern datasets (201 bp). This hyper-variable 201 bp portion of the control region is within fragment 1.

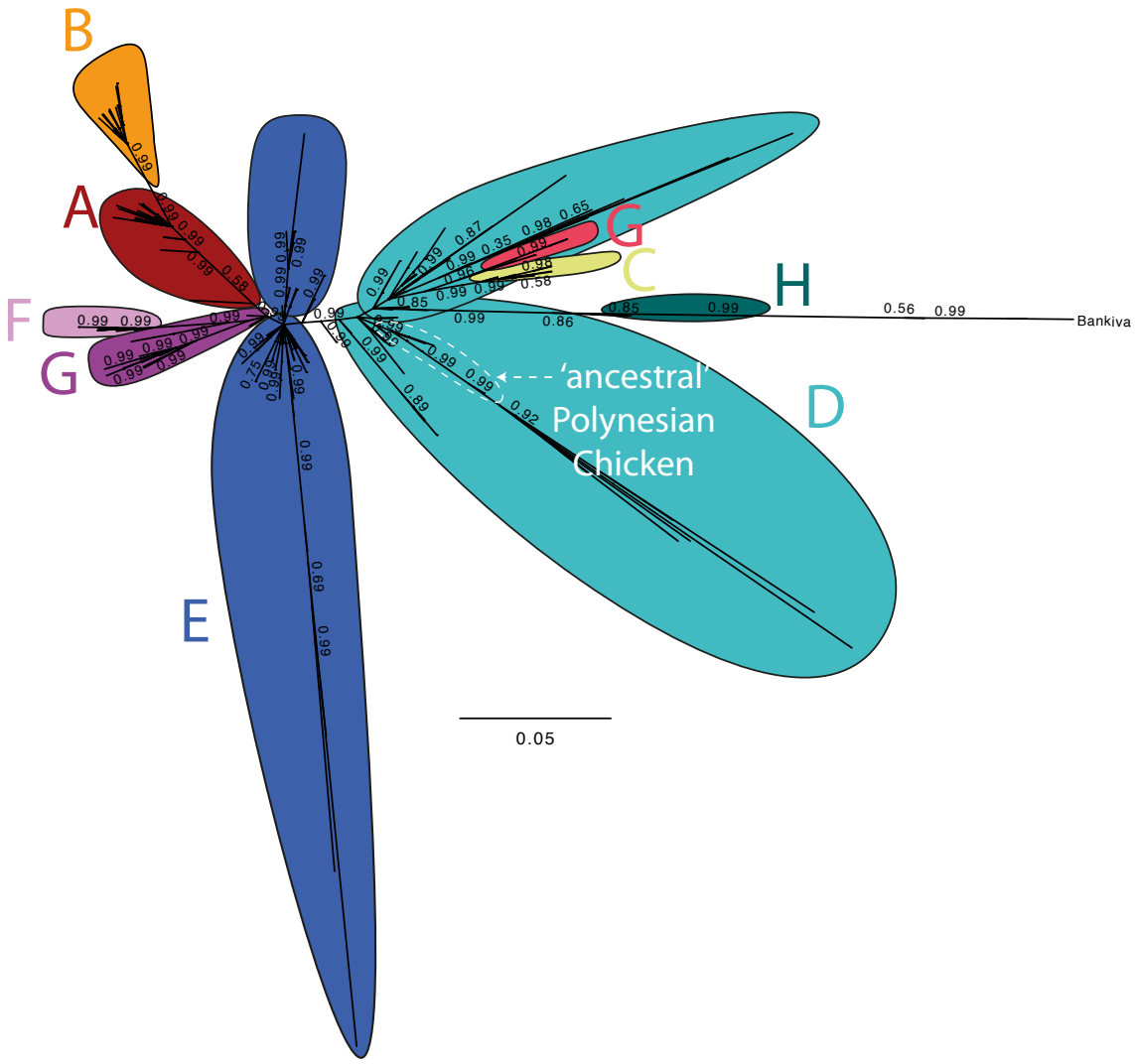


Fig. S14. Maximum Likelihood tree constructed using PhyML with 330bp of mitochondrial control region (mtDNA CR) for 1254 *Gallus gallus* sequences worldwide (see Dataset S6 for list of samples), with *G. g. bankiva* as an outgroup. All 1226 modern sequences used for reference purposes were included, plus the 22 ancient Pacific bone samples and 6 modern feather samples from the Marquesas, as only these were sequenced for the 330 bp fragment of the mtDNA CR. Colors and labels in this figure represent each of the nine worldwide chicken haplogroups initially identified in Liu *et al.* (13), with the addition of our 'ancestral' Polynesian chicken group. The support values on branches are estimated using a $Chi^2$-based approximate Likelihood Ratio Test (aLRT) – the 'ancestral' Polynesian chicken group has branch support of 0.999.

Dataset S1 - Location and dating information for the chicken samples successfully analyzed in this study, plus Storey *et al.*'s (1, 25, 26) and Dancause *et al.* (10) Pacific samples (see http://dx.doi.org/10.6084/m9.figshare.897928).

Dataset S2 – Summary of Pacific samples from this study plus those from Storey *et al*. (1, 25, 26) without the use of Shrimp DNase, and the presence of the ancient haplotypes in modern chicken populations (see http://dx.doi.org/10.6084/m9.figshare.897927).

Dataset S3 – Replication of PAQANA011 - internal (extract ACAD9068 using repeated PCR and Sanger sequencing) and external (at Durham University), and compared to Storey et al.'s (2007) EF535246 (PAQANA011) sequence (see http://dx.doi.org/10.6084/m9.figshare.897929).

Dataset S4 – Cloning results of ACAD internal replication and external replication at Durham (see http://dx.doi.org/10.6084/m9.figshare.897930).

Dataset S5 – Variable sites across all unique haplotypes with number of 'Polynesian' SNPs (columns with PSNPs are highlighted by dark outline) (see http://dx.doi.org/10.6084/m9.figshare.897932).

Dataset S6 - Unique haplotype details of all sequences used in the study (see http://dx.doi.org/10.6084/m9.figshare.897931).

References


1. Storey AA, *et al.* (2007) Radiocarbon and DNA evidence for a pre-Columbian introduction of Polynesian chickens to Chile. *Proc Natl Acad Sci U S A* 104(25):10335-10339.
2. Cooper A & Poinar HN (2000) Ancient DNA: Do it right or not at all. *Science* 289(5482):1139.
3. Pfeiffer I, *et al.* (2004) Forensic DNA-typing of dog hair: DNA-extraction and PCR amplification. *Forensic Sci Int* 141:149-151.
4. Drummond AJ, *et al.* (2011) Geneious v5.4.
5. Lanfear R, *et al.* (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol* 29(6):1695-1701.
6. Ronquist F, *et al.* (2012) MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 61(3):539-542.
7. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22(21):2688-2690.
8. Adebambo A, *et al.* (2010) Lack of phylogeographic structure in Nigerian village chickens revealed by mitochondrial DNA D-loop sequence analysis. *International Journal of Poultry Science* 9(5):503-507.
9. Berthouly-Salazar C, *et al.* (2010) Vietnamese chickens: a gate towards Asian genetic diversity. *BMC genetics* 11:53.
10. Dancause KN, *et al.* (2011) Characterizing genetic diversity of contemporary Pacific chickens using mitochondrial DNA analyses. *PLoS ONE* 6(2):e16843.
11. Gongora J, *et al.* (2008) Indo-European and Asian origins for Chilean and Pacific chickens revealed by mtDNA. *Proc Natl Acad Sci U S A* 105(30):10308-10313.
12. Kanginakudru S, *et al.* (2008) Genetic evidence from Indian red jungle fowl corroborates multiple domestication of modern day chicken. *BMC Evol Biol* 8:174.
13. Liu YP, *et al.* (2006) Multiple maternal origins of chickens: Out of the Asian jungles. *Molecular Phylogenetics and Evolution* 38(1):12-19.
14. Muchadeyi FC, *et al.* (2008) Mitochondrial DNA D-loop sequences suggest a Southeast Asian and Indian origin of Zimbabwean village chickens. *Animal Genetics* 39(6):615-622.
15. Mwacharo JM, *et al.* (2011) Mitochondrial DNA reveals multiple introductions of domestic chicken in East Africa. *Molecular Phylogenetics and Evolution* 58(2):374-382.
16. Oka T, *et al.* (2007) Analysis of mtDNA sequences shows Japanese native chickens have multiple origins. *Animal Genetics* 38(3):287-293.
17. Silva P, *et al.* (2009) Mitochondrial DNA-based analysis of genetic variation and relatedness among Sri Lankan indigenous chickens and the Ceylon junglefowl (*Gallus lafayetti*). *Animal Genetics* 40(1):1-9.

18. Barker FK, *et al.* (2012) Contrasting evolutionary dynamics and information content of the Avian mitochondrial control region and ND2 gene. *PLoS ONE* 7(10):e46403.
19. Langford SM, *et al.* (2013) Australian and Pacific contributions to the genetic diversity of Norfolk Island feral chickens. *BMC genetics* 14:91.
20. Storey AA, *et al.* (2013) Polynesian chickens in the New World: a detailed application of a commensal approach. *Archaeology in Oceania* 48(2):101-119.
21. Guindon S, *et al.* (2010) New algorithms and methods to estimate Maximum-Likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst Biol* 59(3):307-321.
22. Keane TM, *et al.* (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* 6(29):1-17.
23. Huson DH & Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* 23(2):254-267.
24. Bandelt HJ, *et al.* (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16(1):37-48.
25. Storey AA, *et al.* (2012) Investigating the global dispersal of chickens in prehistory using ancient mitochondrial DNA signatures. *PLoS ONE* 7(7):e39171.
26. Storey AA, *et al.* (2010) Mitochondrial DNA from 3000-year old chickens at the Teouma site, Vanuatu. *J Archaeol Sci* 37(10):2459-2468.
27. R development core team (2011) R: A language and environment for statistical computing (R foundation for statistical computing, Retrieved from http://www.r-project.org.).
28. Excoffier L & Lischer H (2010) Arlequin suite ver3.5: A new series of programs to perform population genetics analyses under Linux and Windows. 10:564-567.
29. Anderson CNK, *et al.* (2005) Serial SimCoal: A population genetics model for data from multiple populations and points in time. *Bioinformatics* 21(8):1733-1734.
30. Paabo S, *et al.* (2004) Genetic analyses from ancient DNA. *Annu Rev Genet* 38:645-679.
31. USB (2006) Shrimp DNase, Recombinant. ed Corporation U (USB Corporation, Cleveland), pp 1-2.
32. Leonard J, *et al.* (2007) Animal DNA in PCR reagents plagues ancient DNA research. *J Archaeol Sci* 34(9):1361-1366.
33. Robins JH, *et al.* (2001) Hit or miss? Factors affecting DNA preservation in Pacific archaeological material. *Australasian Connections and New Directions: Proceedings of the 7th Australasian Archaeometry Conference*, eds Jones M & Sheppard P (University of Auckland), pp 295-305.
34. Halloran AF & Halloran AG (1969) The feral chickens (Phasianidae) of French Polynesia. *Proceedings of the Oklahoma Academy of Sciences* 49:169-170.
35. Ellis W (1831) *Polynesian researches: during a residence of nearly eight years in the Society and Sandwich Islands* (Henry G. Bohn, London).
36. Kirch P (2000) *On the Road of the Winds: An Archaeological History of the Pacific Islands Before European Contact* (University of California Press, Berkeley).
37. Intoh M (1996) Multi-regional contacts of prehistoric Fais Islanders in Micronesia. *Indo-Pacific Prehistory: The Chiang Mai Papers*, eds Glover IC & Bellwood B (Australian National University, Canberra), Vol 2, pp 111-118.
38. McNiven IJ, *et al.* (2011) New direction in human colonisation of the Pacific: Lapita settlement of south coast New Guinea. *Australian Archaeology* 72:1-6.
39. Gillespie R, *et al.* (2012) Man and megafauna in Tasmania: closing the gap. *Quat Sci Rev* 37:38-47.
40. McCullagh JSO, *et al.* (2006) Analysis of amino acid C-13 abundance from human and faunal bone collagen using liquid chromatography/isotope ratio mass spectrometry. *Rapid Commun Mass Spectrom* 20(18):2761-2768.
41. Storey AA, *et al.* (2008) Pre-Columbian chickens, dates, isotopes, and mtDNA. *Proc Natl Acad Sci U S A* 105(48):E99.
42. Zhou B, *et al.* (2010) Matrilineal components and genetic relationship of Silkies from China and Japan. *Journal of Poultry Science* 47(1):22-27.
43. Miao YW, *et al.* (2013) Chicken domestication: an updated perspective based on mitochondrial genomes. *Heredity* 110(3):277-282.