

## ARTICLE

# European admixture on the Micronesian island of Kosrae: lessons from complete genetic information

Penelope E Bonnen<sup>\*1</sup>, Jennifer K Lowe<sup>2,3,4</sup>, David M Altshuler<sup>3,4,5,7</sup>, Jan L Breslow<sup>2</sup>, Markus Stoffel<sup>2,9</sup>, Jeffrey M Friedman<sup>2,6</sup> and Itsik Pe'er<sup>8</sup>

The architecture of natural variation present in a contemporary population is a result of multiple population genetic forces, including population bottleneck and expansion, selection, drift, and admixture. We seek to untangle the contribution of admixture to genetic diversity on the Micronesian island of Kosrae. Toward this goal, we used a complete genetic approach by combining a dense genome-wide map of 100 000 single-nucleotide polymorphisms (SNPs) with data from uniparental markers from the mitochondrial genome and the nonrecombining portion of the Y chromosome. These markers were typed in ~3200 individuals from Kosrae, representing 80% of the adult population of the island. We developed novel software that uses SNP data to delineate ancestry for individual segments of the genome. Through this analysis, we determined that 39% of Kosraens have some European ancestry. However, the vast majority of admixed individuals (77%) have European alleles spanning less than 10% of their genomes. Data from uniparental markers show most of this admixture to be male, introduced in the late nineteenth century. Furthermore, pedigree analysis shows that the majority of European admixture on Kosrae is because of the contribution of one individual. This approach shows the benefit of combining information from autosomal and uniparental polymorphisms and provides new methodology for determining ancestry in a population.

*European Journal of Human Genetics* (2010) 18, 309–316; doi:10.1038/ejhg.2009.180; published online 21 October 2009

**Keywords:** admixture; haplotype; population genetics; mitochondria; Y chromosome; Oceania

## INTRODUCTION

Delineating the ethnic background of an individual is important for gene mapping studies, as well as for application of genetic information in a clinical setting. The autosomal genome is a composite of DNA fragments that have been inherited from multiple ancestors and intermingled by meiotic recombination. Complementing these data with the uniparentally inherited mitochondrial genome and Y chromosome provides complete genetic information and offers the ability to explore the full spectrum of genetic variation and history. An accurate determination of ancestry for segments within an individual's genome can be used to understand the genetic history of an individual and the evolutionary origins of a contemporary population.

We investigate the ethnic origins of the contemporary Kosraen population. Kosrae, Federated States of Micronesia is an island, the population of which is estimated at 7700 individuals.<sup>1</sup> As a part of Remote Oceania, Kosrae is situated in the last region of the world to be continuously inhabited by humans. Kosrae was settled just over 2000 years ago.<sup>2</sup> The precise source of settlement of Kosrae remains unclear, possibly by nearby Melanesians; however, there is consensus that the ultimate source for the peopling of this region of the world is continental Asia. In more recent history, Kosrae has experienced contact with individuals from outside the Pacific. Anecdotal evidence

points to the presence of admixture on Kosrae, resulting from the mixing of indigenous individuals with European visitors to the island. Previous studies have noted the presence of European alleles in Polynesia<sup>3–5</sup> and attributed this in various ways to the whalers, sailors, traders, and missionaries who have been present in this region over the past two centuries. However, no such report yet exists for Micronesia.

Data from polymorphisms in the mitochondrial genome and from the nonrecombining portion of the Y chromosome (NRY) have classically been used to track the migrations of humans across the earth. Because of the lack of recombination in these genetic regions, this type of data has the power to reveal the deep evolutionary history of a population and its relationship to other populations. In addition, as the transmission of these markers is strictly maternal or paternal, they bring to bear sex-specific gene flow. The worldwide distribution of specific haplogroups has been mapped for both the mitochondrial genome and the NRY.<sup>6,7</sup> In this study, we use this geographic specificity of haplogroups to shed light on the presence or absence of European alleles on Kosrae.

Admixture is the mixing of individuals from genetically distinct populations. Admixture can be leveraged for gene mapping in diseases that vary in incidence between populations.<sup>8–10</sup> In case-control association studies, admixture can lead to false inflation of test

<sup>1</sup>Department of Human and Molecular Genetics, Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA; <sup>2</sup>Program in Molecular and Human Genetics, The Rockefeller University, New York, NY, USA; <sup>3</sup>Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA; <sup>4</sup>Program in Medical and Population Genetics, The Broad Institute of Harvard and MIT, Cambridge, MA, USA; <sup>5</sup>Center for Human Genetics Research, Massachusetts General Hospital, Boston, MA, USA; <sup>6</sup>Howard Hughes Medical Institute, Chevy Chase, MD, USA; <sup>7</sup>Department of Medicine, Harvard Medical School, Boston, MA, USA; <sup>8</sup>Department of Computer Science, Columbia University, New York, NY, USA

\*Correspondence: Dr PE Bonnen, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, MS BCM 226, Houston, TX 77030, USA. Tel: +1 713 798 4256; Fax: +1 713 798 6977; E-mail: pbonnen@bcm.edu

<sup>9</sup>Current address: Institute of Molecular Systems Biology, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland.

Received 22 May 2009; revised 11 August 2009; accepted 11 September 2009; published online 21 October 2009

statistics when not properly accounted for.<sup>11,12</sup> The ability to determine the locus-specific structure in a population allows for the determination of which loci are affected by substructure and to what extent, hence providing a less conservative and more precise means of correction. We have developed a new software that performs this parsing of population ancestry and applied it to the Kosraen population.

Our software, called Xplorigin, extends the single-marker Hidden Markov Model (HMM) for population ancestry<sup>13</sup> to handle haplotypes rather than single-nucleotide polymorphism (SNP) alleles. This basic HMM was designed to handle a custom panel of unlinked AIMs, and was further extended to the analysis of arbitrary, commercially available sets of markers, taking into account linkage disequilibrium (LD) between successive SNPs.<sup>14–16</sup> This allows researchers to leverage the same data set for delineating ancestry as that for association mapping. As marker density increases, correlations between nonconsecutive SNPs are observed more frequently, and need to be taken into account. We model these explicitly, as we replace SNP alleles by haplotypes and we construct a model of nested HMMs that fully controls for LD between SNPs along the same haplotype to improve resolution of ancestry parsing.

We compiled a complete genetic profile for the majority of the adult population on Kosrae using polymorphism data from the autosomes, mitochondrial genome, and nonrecombining portion of the Y chromosome. We use our software to assess the extent of European and Micronesian ancestry in each individual, and to delineate the actual haplotypes within each person's genome. Data from the mitochondrial genome and NRY complement and extend our findings from the autosomal genome.

## MATERIALS AND METHODS

### Samples: Kosrae and reference populations

Kosrae samples were collected as part of a whole-genome association study for anthropometric and metabolic traits described elsewhere,<sup>17</sup> ascertaining most adults and eventually collecting DNA from 3106 individuals out of 4100 adults on the island.<sup>1</sup> IRB approval was obtained from all participating institutions. From these DNAs, 2988 samples were successfully genotyped and included in these analyses.

We used data from the International HapMap Project CEU (CEPH individuals from European ancestry from Utah), HCB (Han Chinese from Beijing), and JPT (Japanese from Tokyo) samples described in detail elsewhere.<sup>18</sup>

### Kosrae pedigree

Details regarding pedigree construction are available elsewhere.<sup>17</sup> Briefly, 98% of samples belong to a large interconnected pedigree that was initially constructed on the basis of questionnaires and interviews. Relationships in the pedigree have been validated, corrected, or broken apart according to genetic data, culminating in the tightened pedigree used in this study.

### Genotype data

The SNP genotype data presented here were generated using GeneChip Human Mapping 100K SNP arrays by Affymetrix (Santa Clara, CA, USA). Samples and SNPs were filtered as in Lowe *et al*<sup>17</sup> with the exception that we retained monomorphic and low-frequency SNPs that provide vital information regarding population origin. Finally, HapMap data (including monomorphic SNPs) and Kosrae data were filtered to include an intersection of their SNP sets. Ultimately, 106 541 SNPs passed these filters and were used in these analyses.

### Mitochondrial and Y haplogroup typing

Ninety-six males who are at least five meioses unrelated to each other were typed for haplogroup defining mitochondrial (Supplementary Table 1) and Y (Supplementary Table 2) markers. These markers were selected to represent a

worldwide distribution of haplogroups with particular emphasis on distinguishing between East Asian/Oceanic and European haplogroups. Subsequent genotyping was undertaken for all samples in the Kosraen collection for mitochondrial markers M (C10400T, A10398G), M7 (C6455T), B (9 bp deletion 8271–81), P (A15607G), and W (G8994A). Mutations and haplogroup nomenclature were selected from MitoMap<sup>6</sup> and from previous studies of Asian and European mitochondrial genomes.<sup>19–22</sup> All male samples were typed for Y-chromosome markers M216 (Y:C13946958T), M9 (Y: C20189645G), M106 (Y:A20325812G), M175 (Y: 14018105, 5 bp insertion TTCTC), M45 (Y:A20327175G), and M173 (Y:A13535818C).<sup>7,23,24</sup> Haplogroup nomenclature and haplogroup mutation affiliation were used according to the guidelines of the Y Chromosome Consortium.<sup>7</sup> The presence of specific Y haplogroups in Asia and Europe was shown previously.<sup>3,7,20</sup>

### Xplorigin analysis

Xplorigin uses a training data set to approximate haplotype frequencies in the two populations that are the 'source' of admixture. We used HapMap data from 30 CEU trios, and 30 Kosraen trios were used as the training set. Haploview<sup>25</sup> was used to determine LD blocks for both populations separately. We used the SPINE method for determining blocks, as this method includes SNPs that are monomorphic in a population. The two blockfiles were compared and the union set of blocks was determined choosing the smallest common denominator for each haplotype block. Haploview was run again to generate custom 'union' block files for each population. Xplorigin is then run using block files as input for training data and the Kosraen sample data in genotype format.

### Xplorigin software

Formally, we partition the genome into haplotype blocks by any block-partition method of choice, and evaluate haplotypes in each block using training data from each of the source populations. We merge block partitions across different populations, defining a multipopulation block as a set of SNPs in the same block in all source populations. For each population of ancestry  $p$ , we define an HMM  $M(p)$  that has a state for each haplotype in each block, inferring transitions between haplotypes and haplotype content using standard software.<sup>25</sup> Emission probabilities from each haplotype state define the chances of observing each haplotype along a chromosome from population  $p$ , taking into account low-probability genotyping error at each SNP. We further include a 'wildcard' haplotype, representing the potential presence of rare haplotypes in the population, which is unavailable in the training set. The expected population frequency of this undersampled haplotype depends on training set size. Emission probabilities at each SNP reflect its allele frequencies.

For multiple populations,  $P=[\{p_1, p_2, \dots\}]$ , we define an HMM  $M(P)$  for admixed haplotypes with states  $(p_i, h_j)$  for each haplotype  $h_j$  in  $M(p_i)$ . Emissions and state transitions within each population  $p_i$  are copied from  $M(p_i)$ . Transitions between populations have probabilities that depend on recombination rate as the admixture event (Patterson *et al*<sup>13</sup>). This constitutes a model for haplotypes in the admixed population.

Finally, we turn to handle genotype data. Conceptually, we model genotypes by a Cartesian product of two haplotype models. When phased data are available, the emission probability of alleles  $(a_M, a_F)$  at a particular SNP on maternal and paternal chromosomes, respectively, is the product of respective emission probabilities of the two parental HMMs,  $M_M(P)$  and  $M_F(P)$ , at the current SNP. If the phase of a heterozygote genotype is unknown, the emission probability is the sum of the two phase-resolved options for that heterozygote. Our model is thus flexible to optimally handle mixed data from unrelated individuals, as well as from individuals in pedigrees, and is therefore partially phase resolved.

Xplorigin is written in C++ and is available as open source software at <http://www.cs.columbia.edu/~itsik/Xplorigin/Xplorigin.htm>. Software parameters include probability of undersampling a haplotype (see above), the chance of changing the population of ancestry per basepair distance (which is essentially the recombination rate times the number of generations since admixture), and the assumed error rate in the data.

## RESULTS

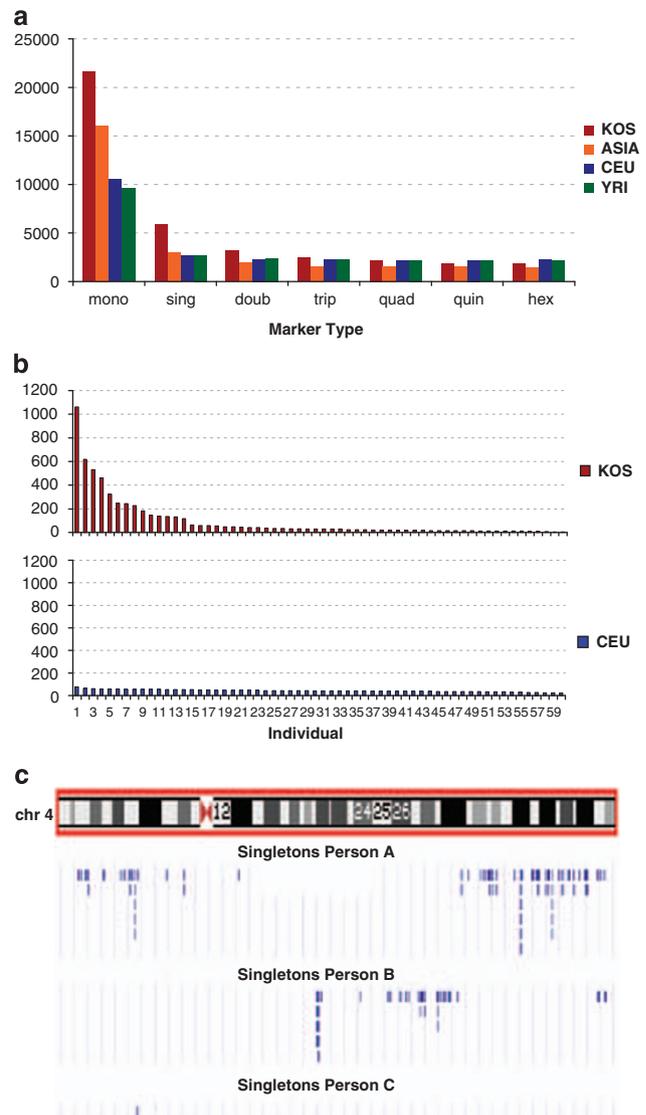
### Evidence for European admixture on Kosrae from single-marker analysis

We previously generated a haplotype map for the Kosraen population by typing ~110 000 SNPs on 30 Kosraen trios.<sup>26</sup> These data showed evidence of European admixture in some individuals. Comparison of haplotypes present in Kosrae *versus* the European, Asian, and African samples used in the HapMap project showed that 3% of haplotypes observed in Kosrae were only found in European HapMap samples, indicating evidence of European admixture in this otherwise Micronesian genetic background (data not shown). Furthermore, the allele frequency spectrum for SNPs in Kosrae showed an increase in the number of low-frequency alleles over what would be expected from a population in equilibrium, and even more so compared with the expectation from a bottlenecked population (Figure 1a). Kosraens had twice as many 'singletons,' SNPs in which the minor allele is observed on only one chromosome, compared with other outbred populations. These singletons clustered within individuals (Figure 1b) and within chromosomal locations (Figure 1c) consistent with the structure of an inherited chromosomal segment. On Kosrae, the number of singletons per individual ranged from 15 to greater than 1000. In contrast, the European trio parents showed a very narrow range in singletons from 25 to 32. These low-frequency alleles are SNPs that are relatively common in Europe, and of low frequency or not in the indigenous Micronesian population (data not shown). Taken together, the singletons in the Kosraen trios seem to be signposts for the presence of European genomic regions within these primarily Micronesian individuals. These data in this representative subset of the population indicate the presence of European admixture on Kosrae. We moved forward and tested our entire cohort to ascertain the extent of admixture on the island.

### Demonstrating efficacy of Xplorigin using individuals of known mixed ancestry

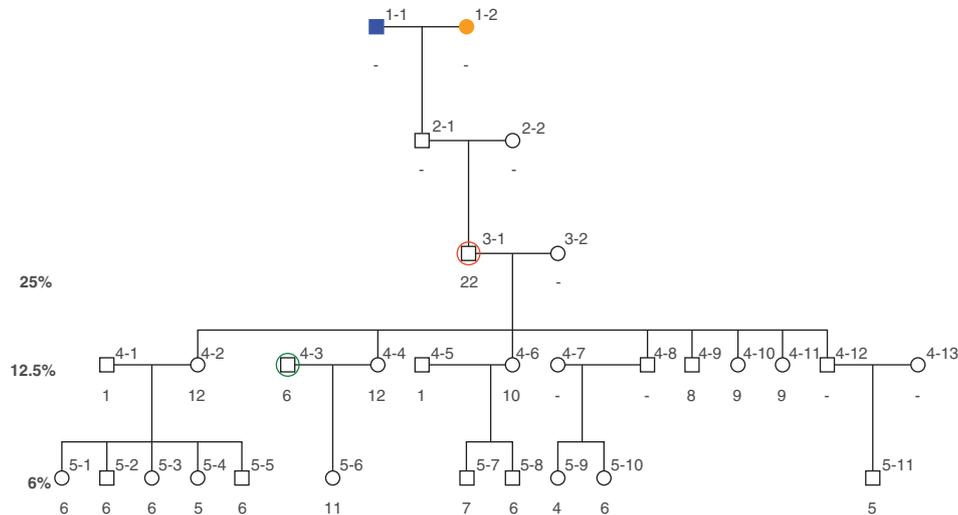
In our software, Xplorigin, the admixture process is described by a top-level HMM that designates population ancestry in each region of the genome. For each ancestral population, we model observed data by a bottom-level HMM with haplotypes as states. This nested HMM model accounts for LD between SNPs and produces a likelihood score that each haplotype belongs to a particular population. Summing the genomic distance spanned by haplotypes demarcated for each population determines the percentage ancestry for each population in an individual genome.

We tested the accuracy of our analysis method for estimating the percentage ancestry for an individual, by using a pedigree of individuals with known ancestry. These individuals are descended from the mating of a European father and a Kosraen mother. Figure 2 shows the portion of this very large pedigree that begins with one grandchild of this mating (labeled 3-1, circled in red). With no additional source of European ancestry besides the grandfather, grandchildren of this mating should be on an average 25% European. Our analysis estimates this grandchild's genome as 22% European, completely in line with expectation. European ancestry is reduced by half in each successive generation (Figure 2). Siblings and first cousins serve as an internal control for each other and show similar amounts of European ancestry. In addition, it is clear from where European ancestry enters the pedigree from additional sources. The person circled in green in Figure 2 (labeled 4-3) married into the family and is estimated to have 6% European ancestry. As a result of this additional contribution of European ancestry, the child of this person (5-6) has higher European ancestry at 11% than her first cousins with 4-7% (5-1 to 5-11).



**Figure 1** Elevation in the number of singletons in Kosrae is a sign of admixture. (a) Increase in the number of low-frequency alleles in Kosrae compared with HapMap population data. 'mono' refers to a monomorph, or SNP in which the minor allele is not observed in a population. 'sing', singleton or SNP in which the minor allele is observed on one chromosome in a population; 'doub' is observed twice; 'trip' is observed thrice; 'quad' is observed four times; 'quin' is observed 5 times. KOS, Kosrae; ASIA, Japanese from Tokyo plus Han Chinese from Beijing; CEPH, North and Western European; YRI, Yoruba from Nigeria. (b) Kosrae singletons cluster within individuals. The number of singletons observed in each trio parent is plotted in rank order. Kosraens are shown in yellow, CEPHs are shown in blue. (c) Kosrae singletons colocalize in physically continuous regions. The top row is a schematic of chromosome 4. Each row below shows the singletons on chromosome 4 for one individual Kosraen. Each blue tick mark represents the presence of a singleton. Person A has a cluster of singletons in the p and q arms of chromosome 4. Person B has a cluster of singletons on 4q. Person C has only one singleton present on chromosome 4 and was subsequently confirmed to not have any European ancestry on chromosome 4.

A summary of the Xplorigin estimate of ancestry by generation for the complete pedigree is shown in Table 1. On an average, the fourth generation of individuals has  $11\% \pm 2.22$  SD European ancestry, the fifth has  $6\% \pm 1.73$  SD, and the sixth generation has  $3\% \pm 1.66$  SD. These estimates by our software are highly accurate compared with the expectation from each individual's position in the pedigree.



**Figure 2** Admixture estimates from Xplorigin agree with known admixed pedigree. The Xplorigin estimate for the percentage of the genome that is European is shown under each person's symbol. Individuals with a '-' were not available for analysis. The expected percentage European ancestry for each generation is shown on the left. The first generation shows a mating between a European male (blue) and a Micronesian female (orange). The third generation shows the grandchild (circled in red) of this mating who is expected to have 25% European haplotypes in their genome. Successive generations show reduced admixture. Sibships and first cousins share similar amounts of admixture. An individual who contributed an additional source of European ancestry into the pedigree is identified by Xplorigin (circled in green).

**Table 1** Evaluation of Xplorigin using pedigree of known mixed ancestry

Generation post-admixture	Direct descendants (n=810)	Individuals in pedigree with data (n=629)	Individuals in pedigree with data and 1 source of European (n=468)	Mean	Median	SD
2	2	0	0			
3	21	2	2	23.00	23	2.83
4	126	68	61	10.69	11	2.22
5	419	343	267	5.66	6	1.73
6	232	207	129	3.14	3	1.66
7	10	9	9	1.11	1	0.78

Evidence from mitochondrial, Y chromosome, autosome, name, and anecdotal data concludes that the mating between a European man and a Micronesian woman resulted in a pedigree of 810 direct descendants. We used this pedigree to test our software. For each generation of the pedigree the mean, median, and SD of the amount of European ancestry is shown. There are places in the pedigree where European ancestry enters from additional sources. These calculations were completed using only the individuals in the pedigree whose European ancestry derives only from the original European founder.

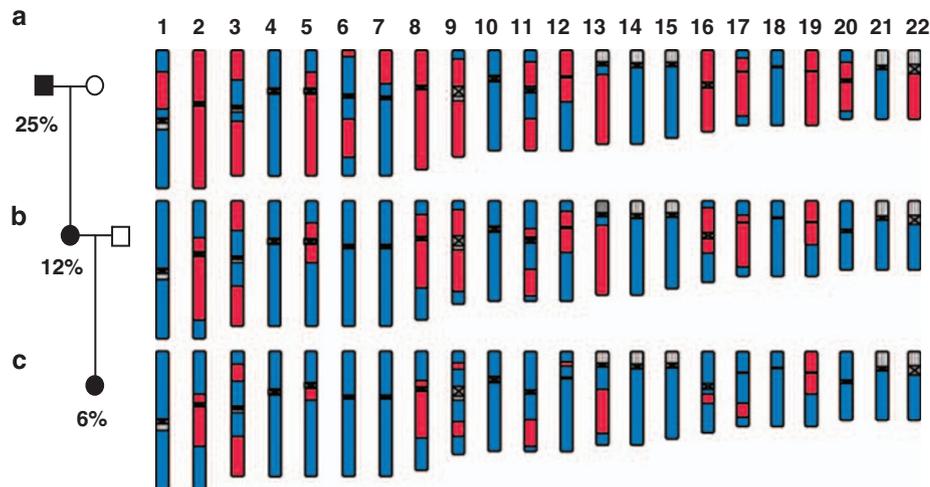
We examined the inheritance of European haplotypes observed in this pedigree (Figure 3). Meiosis in the child of an admixed mating is the first opportunity to generate admixed chromosomes. Furthermore, it is the only meiosis in which all crossovers result in a change in ancestry along the chromosome. We examined the Xplorigin designation for haplotype ancestry throughout the genome of the grandson of the admixed mating shown in Figure 2 (individual 3-1). We noted crossover events as the junction between European and Micronesian haplotypes in this individual. We observed 19 crossovers in this male meiosis, which is within the range observed for male meioses in other studies.<sup>27</sup> Not all chromosomes experienced a crossover and, as such, 11 chromosomes are either completely European or Micronesian. The chromosomes that are admixed most often experienced recombination near the telomeres as reported in other studies.<sup>27,28</sup>

Using the definition of centiMorgan as the guide for the length of haplotypes, after  $k$  post-admixture meioses, the length of European haplotypes should be  $100 \text{ cM}/k$ . A grandchild of an admixed mating is  $k=1$ . There was a fairly wide range in the length of European haplotypes in the grandchild we studied (individual 3-1), 34–230 Mb/17–197 cM, with an average length of 77 Mb/68 cM.

However, recalculating the length of European segments from only the larger chromosomes yields an average length of 96 cM, which is in close agreement with prediction. A further study of individuals in this pedigree showed that the length of European haplotypes decreased with each generation as expected. The greatgrandchild (4-2) has an average European haplotype length of 64 Mb/41 cM (range 13–147 Mb/4–77 cM). The next generation (5-1) is  $k=3$  and has an average length of 39 Mb/34 cM (range 5–100 Mb/3–121 cM), which is in line with theoretical expectation.

#### Genome-wide map of ancestry delineated for each Kosraen

We estimated genome-wide ancestry for each Kosraen ( $N=2988$ ) individually using the 106 541 SNP data set (see Materials and Methods section). The majority of Kosraens show no evidence for the presence of European admixture. As shown in Figure 4a, 62% of Kosraens have 0–1% European ancestry. Twenty-eight percent of Kosraens have European ancestry ranging from 2 to 10%. Ten percent have European ancestry higher than 10% and just 1% of Kosraens have European ancestry greater than 25%. The generally low admixture rates at the population level allow association analysis to disregard



**Figure 3** Inheritance of European haplotypes. The inheritance of European segments of the genome through two generations. Each of the three rows of the figure shows autosomes for one individual in the pedigree. The first row shows an individual (a), who is the grandchild of a mating between a European and a Kosraen, and therefore is expected to have 25% European contribution to their genome. (b) The second row is the child of person (a) and a Micronesian individual, and therefore is expected to be 12.5% European. (c) The person represented in the third row is the offspring of person (b) and a Micronesian and thus has 6% European genome. Chromosomes are painted red where the Xplororigin analysis indicates European segments and blue where there are Micronesian segments. The chromosomal segments shown in gray were not ascertained by the set of SNPs used in this study. Because there is no other European contribution to this pedigree, recombination events can be observed when a European segment is shorter in the successive generation.

European contribution and use admixture-agnostic methods to search for causal genetic factors.<sup>17</sup>

#### Timing and source of introduction of European ancestry on Kosrae

To understand the source and timing of the introduction of admixture on Kosrae, we genotyped all individuals in our study population for mitochondrial and Y-chromosome polymorphisms that distinguish between Asian and European populations. Out of the 3058 individuals typed, only one Kosraen showed a European maternal lineage, haplogroup W (see below). The maternal lineage of all other samples was Asian. The vast majority of individuals (68%) possessed mitochondrial haplogroup B, 31% had haplogroup M, and 1% had P (Figure 5a). These haplogroups are commonly observed throughout Asia and Oceania.<sup>19–21</sup> Data from the nonrecombining portion of the Y chromosome indicated that most Kosraen men are descended from an Asian paternal lineage. Forty-seven percent of the Kosraen men typed were haplogroup K (xM, O, PQR) (Figure 5b). The next most common Y haplogroups on the island are M and O, with 19 and 18% of men, respectively. Haplogroup C was also observed at a frequency of 12%. All of these haplogroups are observed in Asia and Oceania.<sup>3,7,20</sup>

Three percent of Kosraen men have a European-specific haplotype, R1, on their Y chromosome, indicating paternal European ancestry in these individuals. An examination of the individuals who possess a European Y chromosome reveals an extended pedigree with a male European founder who arrived on the island in the mid-1800s. This founder has 810 direct descendants, 686 of whom are members of our study collection. Our genetic data, both autosomal and mitochondrial, indicate that the mother of the person's children was Micronesian. Fifty-seven percent of individuals with European ancestry in our cohort are descended from this mating.

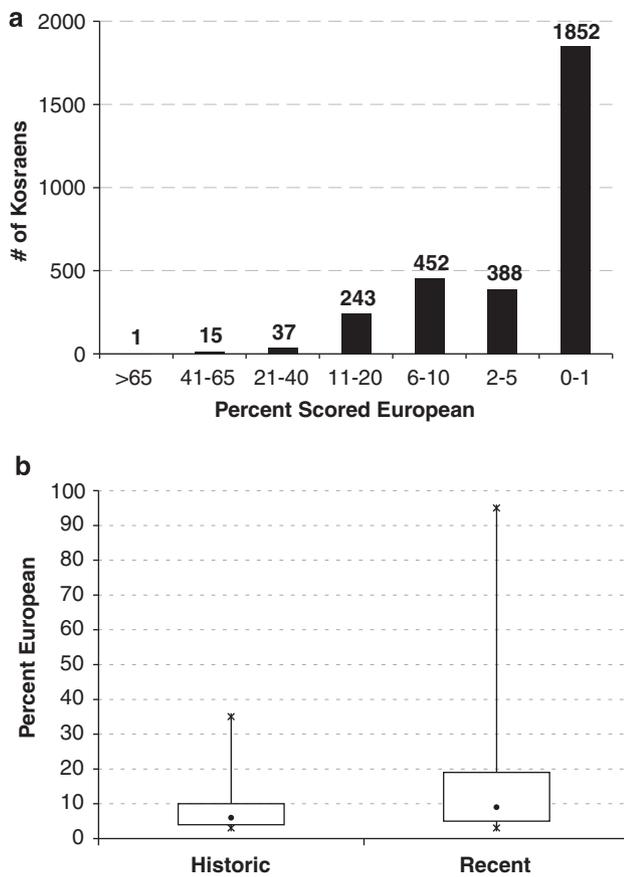
One additional pedigree of 55 individuals carried the European Y haplogroup, R1; however, most pedigrees that showed European admixture did not show a European Y chromosome or European mitochondrial haplogroup (Table 2). This is, of course, possible under various circumstances, including if the European founder was male and had only female children or conversely if the European founder

was female and had only male children. From the pedigree information, we were able to discern when a pedigree must have had a male European founder as well as their estimated date of arrival on the island. All of the larger, admixed pedigrees that we examined had European founders who contributed to the Kosraen gene pool in the late 1800s to early 1900s (Table 2). Together, these seven pedigrees account for 62% of the admixed individuals on Kosrae. We refer to the individuals who are descended from this type of founders as having 'historical admixture.'

The majority of individuals on Kosrae who show European admixture inherited their European genes from these older events. However, in general, these subjects also have a lesser percentage of their genome as being European. No Kosraens with 'historical admixture' have a European contribution to their genome greater than 35% (Figure 4b). All individuals with European admixture greater than 35% are recent migrants to the island and have not been on the island for more than one generation (Figure 4b). There were 10 men living in Kosrae who possess the Y haplotype R1 and who were recent immigrants to the island. All of these men arrived on the island in the late twentieth century and most had only one blood relative in our collection. The single individual who showed the highest amount of European contribution to their genome was 95% or basically all European. This individual is also the only person on the island who had a European mitochondrial haplogroup. This subject is the father of one child and has no other blood relatives reported. There are ~100 individuals in our Kosraen collection who are admixed and who have not been on the island for more than two generations. These individuals contribute 'recent admixture' to the population of the island.

#### DISCUSSION

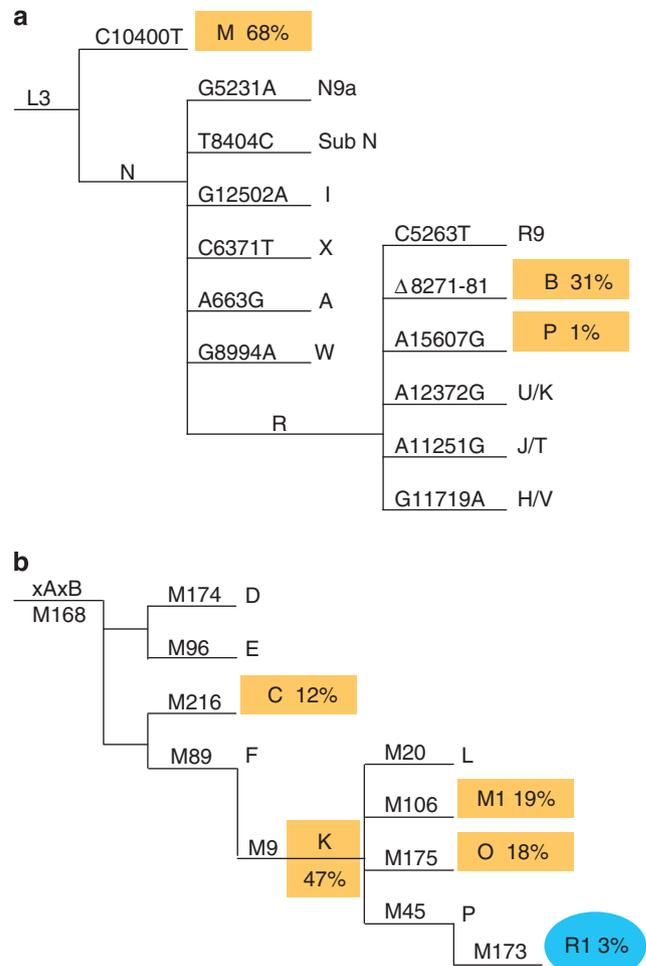
We present analysis of admixture in an entire population using genetic information from nuclear and mitochondrial genomes. We developed new software that enabled quantification of the real contribution of nonindigenous variation to the island population. Delineating admixed ancestry for each sample in this cohort and for each segment



**Figure 4** Extent of European ancestry in 2988 Kosraen genomes. (a) The percentage of genome with European ancestry. The percentage of each individual's genome that is European was calculated using Xplorigin. This percentage for each individual analyzed was plotted into bins. (b) Historical versus recent admixture. A total of 1032 Kosraens have European segments spanning 3% of their genome. These individuals were divided into two categories of recent or historical admixture, depending on when the initial admixture event occurred between the 1800s and early 1900s ( $N=923$ ) or more recently in the past 50 years ( $N=109$ ). Individuals who have recent admixture belong to families that have been present on the island for no more than two generations, whereas individuals who have historical admixture have a family history on the island greater than two generations. This box plot shows the percentage genome that is European for Kosraens falling into these two categories. The extent of the genome that is European is significantly greater in the more recently admixed individuals (Mann-Whitney two-sided  $P$ -value  $6.77E-6$ ).

along the genome allows us to distinguish two distinct waves of introduction of nonindigenous genetics to Kosrae, historical admixture from the late nineteenth to early twentieth century versus recent admixture of first- and second-generation immigrants. Our data from the Y chromosome and mitochondrial genome confirm that the 'historical admixture' was male introduced. Records of whaling ships visiting the island further promote the assumption that historical admixture, even in pedigrees in which we do not observe a European Y chromosome, was most likely derived from men. Other studies using uniparental markers observed primarily male-mediated European gene flow into Polynesian populations,<sup>3-5</sup> but this is the first report for Micronesia.

Despite the large number of Kosraens showing admixture as a result of the first wave of immigration to the island, the overall extent of European ancestry in each individual is relatively low. This is primarily because of the rapid population expansion in Kosrae over the last



**Figure 5** Mitochondrial and Y chromosome haplogroups present in Kosrae. Schematic of the relationships among the mitochondrial (a) and Y chromosome (b) haplogroups tested in Kosrae samples. Haplogroups found to be present in Kosraen samples are shown with the percentage of Kosraens who possess the haplogroup. Haplogroups that are present in Kosrae have been color coded to indicate whether each haplogroup would be expected to be observed in a European (blue) or East Asian/Oceanic (orange) population.

**Table 2** The largest admixed pedigrees detected on Kosrae

Founder/s	No. of direct descendants	European mt/Y	% European	Estimated date of arrival on Kosrae <sup>a</sup>
1	810	Eur Y	100	1870
2	55	Eur Y	100	1910
3	45	No	50	1905
4	40	No	50	1905
5, 6 <sup>b</sup>	32	No	50/25 <sup>c</sup>	1895
7, 8 <sup>b</sup>	28	No	50/25 <sup>c</sup>	1905
9	21	No	100	1900

<sup>a</sup>Date of arrival is estimated based on year of birth of first child or grandchild on record.

<sup>b</sup>Founder pair. One or both members of the pair may be European.

<sup>c</sup>If one member of the pair is European that person is 50% European. If both members are European, then each one is 25% European.

century, as is evidenced by the large pedigrees of admixed individuals now present on the island. In addition, these large pedigrees typically have European ancestry coming from a single, male founder. Owing to the low level of admixture at the population level, efforts to map loci

affecting phenotypic variation using admixture mapping would be severely underpowered.

This demographic history provides us with the opportunity to follow up European haplotypes of individual founders throughout their large pedigrees and make interesting observations on meiotic recombination rates. We observe 19 meiotic crossovers in one male individual, which is in line with observations by Cheung *et al.*<sup>27</sup> We measured the length of genomic segments between recombination events. For the third generation after admixture, we observe European haplotype lengths that fit to theoretical expectation for distance between recombination events (34 cM). For the first and second generations, we observe an average European haplotype length of 68 and 41 cM, respectively, somewhat shorter than the expectation based solely on recombination. However, when only larger chromosomes are considered, the average length for  $k=1$  is 96 cM and for  $k=2$  is 50 cM, perfectly in accord with expectation.

The admixture on Kosrae was introduced relatively recently compared with other well-studied genetically admixed populations. In addition, in Kosrae, most admixture can be attributed to a relatively small number of European founders introduced into the population 5–7 generations ago. This demography results in an unusual finding of increased low-frequency alleles compared with other outbred, populations. Low-frequency alleles are typically considered as new mutations. However, in this case, they are a sign of admixture. The singletons are alleles that did not exist in the indigenous population and were brought in by Europeans. Most of these SNPs have population frequencies in the ancestral populations that reflect the signature of Ancestry Informative Markers: increased frequency in European population and low frequency in Asian population (data not shown). However, some are present in both the HapMap Asian and European populations in relatively equal frequency. The fact that these particular alleles appear as low frequency in Kosrae is most likely because of the introduction of new alleles after the bottleneck. We have not detected significant amounts of other sources of immigration to the island.

On a broader view, the cohort and data in this study provide a unique example of near complete population genetics: examination of whole-genome variation in autosomes, mitochondrial genome, and Y-chromosome markers in essentially every family on the island. It is noteworthy that some admixed pedigrees show neither a European Y nor a European mitochondrial haplogroup, resulting in ambiguity with regard to the sexual origin of this admixture. In studies that only include genetic information from uniparental markers, the European contribution to the ancestry of these individuals would be completely missed. In fact, if our study only examined mitochondrial and Y chromosomes, we would have detected European ancestry in 3% of the population, instead of the 39% we found using a complete genetic approach. Although autosomal data provided the most complete assessment of population history with regard to admixture, the information contained in the uniparental markers is not superseded. Owing to the absence of recombination in the mitochondrial genome and to the nonrecombining portion of the Y chromosome, valuable information with regard to population history is retained. These genetic segments reveal the Kosraen population history vastly deeper into the past than do autosomes, and of course reveal the sex-specific contribution to the contemporary gene pool. Studies of Polynesian<sup>29</sup> and Melanesian<sup>30</sup> populations using serum protein markers and HLA typing have observed the presence of European admixture, but these studies did not have genome-wide resolution and could easily have underestimated the extent of admixture. Owing to the density of our whole-genome data and to combining these data with sex-specific

markers, we are better able to quantify the actual extent of European alleles in a population. These data coupled with the unique population structure in Kosrae allowed us to date with high precision the introduction of European founders to the island. Our results using a genetic approach match historical record, which dates Kosrae's first contact with Westerners to the 1820s.<sup>2</sup> Our methodology serves as a model for studies that seek to delineate and quantify ancestry in any population.

We view this approach as a model for a near future reality of such information available for numerous, larger populations, with dynamics of heredity being observed in unprecedented detail. Although we show a successful parsing of per-locus ancestry at the individual level, this is assisted by population isolation and separation between source populations. For more challenging cohorts with genetically closer founder populations, our analysis brings forth a way to better tease apart admixed segments.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

We acknowledge the invaluable contribution of Vita Skilling, Secretary of the Department of Health and Social Affairs for the Federated States of Micronesia, and thank her for her long-standing collaboration, without which this project would not be possible. We acknowledge Hiroshi Ismael who supported our studies in his capacity as a public servant and by dedicating himself to the health of the people of his homeland. We are deeply grateful to the people of Kosrae and to the Kosrae Department of Health Services for making this study possible. We thank Maude Blundell and Jackie Salit for their diligent work in creating and curating the pedigree. IP was partially supported by NIH 5 U54 CA121852.

- 1 Federated States of Micronesia National Statistics Office: 2000 Census of the Federated States of Micronesia, Palikir, Pohnpei, 2000.
- 2 Segal HG: *Kosrae: The sleeping lady awakens*. Kosrae State Tourist Division, Department of Conservation and Development, Kosrae State Govt: Federated States of Micronesia, 1995.
- 3 Underhill PA, Passarino G, Lin AA *et al*: Maori origins, Y-chromosome haplotypes and implications for human history in the Pacific. *Hum Mutat* 2001; **17**: 271–280.
- 4 Hurles ME, Irven C, Nicholson J *et al*: European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. *Am J Hum Genet* 1998; **63**: 1793–1806.
- 5 Ghiani ME, Moral P, Mitchell RJ *et al*: Y-chromosome-specific STR haplotype data on the Rapanui population (Easter Island). *Hum Biol* 2006; **78**: 565–578.
- 6 MITOMAP: A Human Mitochondrial Genome Database, 2008.
- 7 Karafet TM, Mendez FL, Meilerman MB *et al*: New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res* 2008; **18**: 830–838.
- 8 Chakraborty R, Weiss KM: Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc Natl Acad Sci USA* 1988; **85**: 9119–9123.
- 9 Rife DC: Populations of hybrid origin as source material for the detection of linkage. *Am J Hum Genet* 1954; **6**: 26–33.
- 10 Smith MW, O'Brien SJ: Mapping by admixture linkage disequilibrium: advances, limitations and guidelines. *Nat Rev Genet* 2005; **6**: 623–632.
- 11 Freedman ML, Reich D, Penney KL *et al*: Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004; **36**: 388–393.
- 12 Marchini J, Cardon LR, Phillips MS *et al*: The effects of human population structure on large genetic association studies. *Nat Genet* 2004; **36**: 512–517.
- 13 Patterson N, Hattangadi N, Lane B *et al*: Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004; **74**: 979–1000.
- 14 Sankararaman S, Sridhar S, Kimmel G *et al*: Estimating local ancestry in admixed populations. *Am J Hum Genet* 2008; **82**: 290–303.
- 15 Tang H, Coram M, Wang P *et al*: Reconstructing genetic ancestry blocks in admixed individuals. *Am J Hum Genet* 2006; **79**: 1–12.
- 16 Sundquist A, Fratkin E, Do CB *et al*: Effect of genetic divergence in identifying ancestral origin using HAPAA. *Genome Res* 2008; **18**: 676–682.
- 17 Lowe JK, Maller JB, Pe'er I. *et al*: Genome-wide association studies in an isolated founder population from the Pacific Island of Kosrae. *PLoS Genet* 2009; **5**: e1000365.

- 18 Altshuler D, Brooks LD, Chakravarti A *et al*: A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.
- 19 Kivisild T, Tolk HV, Parik J *et al*: The emerging limbs and twigs of the East Asian mtDNA tree. *Mol Biol Evol* 2002; **19**: 1737–1751.
- 20 Kayser M, Brauer S, Cordaux R *et al*: Melanesian and Asian origins of Polynesians: mtDNA and Y chromosome gradients across the Pacific. *Mol Biol Evol* 2006; **23**: 2234–2244.
- 21 Pierson MJ, Martinez-Arias R, Holland BR *et al*: Deciphering past human population movements in Oceania: provably optimal trees of 127 mtDNA genomes. *Mol Biol Evol* 2006; **23**: 1966–1975.
- 22 Rieder MJ, Taylor SL, Tobe VO *et al*: Automating the identification of DNA variations using quality-based fluorescence re-sequencing: analysis of the human mitochondrial genome. *Nucleic Acids Res* 1998; **26**: 967–973.
- 23 Underhill PA, Passarino G, Lin AA *et al*: The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. *Ann Hum Genet* 2001; **65**: 43–62.
- 24 Underhill PA, Jin L, Lin AA *et al*: Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res* 1997; **7**: 996–1005.
- 25 Barrett JC, Fry B, Maller J *et al*: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- 26 Bonnen PE, Pe'er I, Plenge RM *et al*: Evaluating potential for whole-genome studies in Kosrae, an isolated population in Micronesia. *Nat Genet* 2006; **38**: 214–217.
- 27 Cheung VG, Burdick JT, Hirschmann D *et al*: Polymorphic variation in human meiotic recombination. *Am J Hum Genet* 2007; **80**: 526–530.
- 28 Kong A, Barnard J, Gudbjartsson DF *et al*: Recombination rate and reproductive success in humans. *Nat Genet* 2004; **36**: 1203–1206.
- 29 Deka R, Mc Garvey ST, Ferrell RE *et al*: Genetic characterization of American and Western Samoans. *Hum Biol* 1994; **66**: 805–822.
- 30 Trejaut J, Dunckley H, Sullivan J *et al*: Analysis of HLA class II allogenotyping in Australian aborigines and Papua New Guinea populations. *Tissue Antigens* 1992; **39**: 1–7.



This work is licensed under the Creative Commons Attribution-NonCommercial-No Derivative Works 3.0 Licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/3.0/>

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)