Int J Legal Med (2002) 116 : 99–108

ORIGINAL ARTICLE

Peter Forster · Francesco Calì · Arne Röhl
Ene Metspalu · Rosalba D'Anna · Mario Mirisola
Giacomo De Leo · Anna Flugy · Alfredo Salerno
Giovanni Ayala · Anastasia Kouvatsi · Richard Villems
Valentino Romano

Continental and subcontinental distributions of mtDNA control region types

Received: 30 April 2001 / Accepted: 20 August 2001

Abstract When the mtDNA profile of a crime scene matches that of a suspect, it is necessary to determine the probability of a chance match by consulting the frequencies of the identified allele in a "reference population". The ceiling principle suggests that that population should be chosen in which the allele of the suspect is found at the highest frequency, in order to give the suspect the maximum benefit of doubt. Recently, we advocated the use of a worldwide mitochondrial database combined with a geographical information system to identify the regions of the world with the highest frequencies of matching mtDNA types. Here, we demonstrate that the alternative approach of defining a ceiling reference population on the basis of continent or phenotype (race) is too coarse for a non-negligible percentage of mtDNA control region types.

Keywords DNA fingerprint · Sicily · Greece · Ethnic · Maternal descent

The authors P.F. and V.R. contributed equally to the present work.

P. Forster (✉)
The McDonald Institute for Archaeological Research,
University of Cambridge,
Downing Street, Cambridge CB2 3ER, England
e-mail: pf223@cam.ac.uk,
Fax: +44-1223-339285

F. Calì · G. Ayala · V. Romano
Laboratorio di Genetica Molecolare, Istituto OASI (I.R.C.C.S.),
Troina, Italy

A. Röhl
Institute of Legal Medicine, University of Münster, Germany

E. Metspalu · R. Villems
Estonian Biocentre and Dept. of Evolutionary Biology,
Tartu University, Tartu, Estonia

R. D'Anna · M. Mirisola · G. De Leo · A. Flugy · A. Salerno
V. Romano
Dip. Biopatologia e Metodologie Biomediche,
University of Palermo, Palermo, Italy

A. Kouvatsi
Dept. of Genetics, Aristotle University, Thessaloniki, Greece

Introduction

Mitochondrial DNA control region typing often is the last resort in forensic and ancient DNA cases where the amount of intact DNA may be greatly reduced. The high copy number of mtDNA ensures a better PCR amplification success rate over nuclear loci, but mtDNA inherently has a serious disadvantage: mitochondrial profiles are not unique but shared by many maternally related humans unless mutations have occurred. Therefore, when the mtDNA profile of a crime scene or in a maternity case matches that of a suspect or the putative parent, it is especially important to determine the probability of a chance match by consulting the frequencies of the identified alleles in a "reference population". According to the ceiling principle (National Research Council Committee 1992), in general that reference population should be chosen in which the allele of the suspect is found at highest frequency in order to give the suspect the maximum benefit of doubt. We have previously argued (Röhl et al. 2001) that a worldwide database search of exact and closest genetic matches to a given sequence can assist in the choice of a geographically defined reference population, which incidentally may or may not correspond to populations defined by culture (language) or phenotype (race). Thus our approach differs from the usual practice of basing the reference population database on language or ethnic group (Wittig et al. 2000) or phenotype (e.g. Melton et al. 2001). In contrast to a DNA profile based on multiple independently inherited loci (where the recommendations published in NRC 1996 apply), any single locus such as mtDNA can immediately lose any relationship between phenotype and genotype when admixture has occurred in the maternal line at any time in the past. Secondly, even if admixture can be discounted, we would expect a pooling of mtDNA types into Caucasoids, Hispanics etc., to be too coarse in many instances. This is because evolutionarily

young alleles would have had less time to spread than evolutionarily older alleles. For example, an Italian suspect whose mtDNA matches the crime stain might have an mtDNA type created by mutation only a few thousand years ago in Italy and consequently rare in the rest of Europe. In this case, the chance matching probability determined from a general Caucasoid database would appear to strengthen the case against him. We proposed instead (Röhl et al. 2001) that the choice of the reference population should be geographic and should be based on the genotype actually identified in each case (cf. NRC 1992), or if exact matches are absent in the database, on close matches to the genotype.

It has been argued, particularly in the case of European DNA, that there is no significant subcontinental geographic structure in mtDNA control region sequences (Pult et al. 1994; Cavalli-Sforza and Minch 1997; Simoni et al. 2000; but see Helgason et al. 2001) and the widespread occurrence of even specifically European branches of the evolutionary mtDNA tree (e.g. haplogroup V in Torroni et al. 1998) might appear to support this contention. However, the former studies are based on summary statistics of entire samples and the latter study is intentionally based on a phylogenetic grouping of different control region sequences. Neither approach therefore directly answers the question of how many mtDNA control region types are expected to have a local distribution. To address this issue, we present here a database tool within the mtDNA database "mtradius" which attempts to distinguish geographically widespread sequence types from more localised ones. In this procedure, an mtDNA sequence is entered, for which the most similar genotypes are searched, the identified best matches are then geographically summarised by a trigonometrically determined centre of gravity of genotype frequencies, the database tool then attempts to quantify the geographic dispersion using the standard deviation associated with the centre of gravity. We explore the properties of the method with mtDNA sequences from 100 individuals randomly drawn worldwide and then test the conclusions by applying the method to a German village sample of 1,123 individuals (Pfeiffer et al. 2001) and to a new sample of 159 west Sicilians as a negative control for the Germans.

Subjects and methods

Terminology

In the forensic and anthropological literature, the terms "race", "ethnic group" and "population" are used with conflicting meanings. To avoid confusion, we adhere to the following definitions: "race" signifies a group of individuals sharing a set of heritable phenotypic traits (Nature Genetics editorial 2000); the term is not defined by genetic diversity (Lewontin 1972); "ethnic group" signifies a group of individuals who identify themselves as members of one group (Weber 1922), e.g. on the basis of language, culture, race or religion; the term is not a euphemism for "race". "Population" signifies a group of individuals at one moment in time sharing at least one characteristic (e.g. place of residence) defined by the researcher; it is neither a euphemism for "race", nor do we necessarily imply a unit of genetic continuity through time or a randomly mating endogamous unit, though this narrower definition is commonly used in theoretical population genetics.

Samples and sequencing

For the database analyses, the "mtradius" database (Röhl et al. 2001) was supplemented with new mtDNA sequences from 473 Sicilians (including 106 sequences from Castellamare published by Calì et al. 2001) and 83 Greeks sequenced in Troina and in Thessaloniki, as well as 522 Sicilians sequenced in Tartu. For the sequencing in the Troina laboratory, blood samples were obtained from healthy blood donors of both sexes, selected for third-generation maternal ancestry from the towns of Piazza Armerina (n = 42), Sciacca (n = 80), Troina (n = 100), Caccamo (n = 56), Ragusa (n = 56), and Butera (n = 33). Blood samples from 83 Greeks whose ancestry was traced for three generations were obtained from the following regions: Argolis (n = 15), east Crete (n = 10), east Macedonia (n = 13), Euboea (n = 15), Chios (n = 15) and Lakonia (n = 15).

Samples for sequencing in the Tartu laboratory were obtained from donors who had traced their maternal origins to diverse locations within the following provinces: Agrigento (n = 79), Catania (n = 81), Caltanissetta (n = 69), Enna (n = 4), Messina (n = 61), Palermo (n = 76), Ragusa (n = 3), Siracusa (n = 83) and Trapani (n = 52). A further 10 individuals originated from mainland Italy. Two of Sicily's nine provinces, underrepresented in the Tartu data (Ragusa and Enna provinces), are represented by the towns of Troina (province of Enna) and Ragusa (province of Ragusa) in the data from the Troina laboratory, providing a complete geographic coverage of the island.

The numbering of nucleotide positions follows that of Anderson et al. (1981). Sequencing of the mtDNA control region from np16023 to at least np16391 was performed in the Troina laboratory as previously described (Calì et al. 2001), and in the region from np16024 to np16392 by the Tartu laboratory according to the standard protocol of Amersham-Pharmacia MegaBace and ABI 377 (Perkin Elmer). The sequencing results for the samples from Agrigento province, Sciacca, and Greece are shown in Table 1.

Database functions

The geographic information system "mtradius" identifies the closest matches to a given sequence and displays the geographic distribution of the closest matches on a world map as a centre of gravity with a standard deviation. The current database (a previous version was described by Röhl et al. 2001) comprises mtDNA control region sequences of 17,332 individuals, including the 1,064 Sicilians and Greeks sequenced for this study. Those sequences that were taken from the literature were extensively researched to identify and eliminate inadvertently published duplications and errors as well as non-representative sequences (e.g. intentionally preselected by haplogroup status). The majority of individuals, namely 14,812, were sequenced in the range from nucleotide position (np) 16093 to np16362 (numbering as in Anderson et al. 1981) in the hypervariable region 1 (HVR1). Hence we chose this range for determining the geographic centres of gravity for each mtDNA sequence. Of the 14,812 individuals we classified 11,893 individuals to be representative, with adequate geographic documentation and with less than 6 ambiguous nucleotides within the sequence. Our aim is to distinguish geographically widespread sequences from more localised ones. The following procedure defines a centre of gravity for which a dispersion (standard deviation) is calculated. The centre of gravity, i.e. the weighted average geographical location (expressed as a longitude and a latitude) of a set of best matches is calculated as follows: the mtDNA sequence database is first reduced to an operational database by specifying sequence length and sequence quality, the latter identifiable by annotation and by the number of ambiguous nucleotides. The operational

database contains a finite set X of taxa x with frequencies f_x, coordinates longitude $lon_x \in [-\pi, \pi]$ and latitude $lat_x \in [-\pi/2, \pi/2]$ and a set C_x of character states differing from the published sequence of Anderson et al. (1981). The queried sequence s differs in characters C_s from Anderson et al. (1981). Each character c from the union C of all sets C_x, $x \in X$ is weighted with weight w_c which is set by default to 1 for all characters except for certain inconsistently scored length polymorphisms, which are set to zero by default (Röhl et al. 2001).

The set M of best matches within the database is calculated as follows:

$$M = \{x \in X | d(x,s) = Min\{d(x,s) | x \in X\}\} \text{ where } d(x,s) = \sum_{\substack{c \in C_s \cup C_x \wedge \\ c \notin C_s \cap C_x}} w_c$$

Put simply, the set M is the set of taxa within the database having the minimum distance to the sequence s where the distance between two sequences is calculated as the sum of weights of different character states.

To calculate the centre of gravity, all taxa within the set M are transformed into vectors within the three-dimensional Euclidian vector space, taking radius 1 (as the Earth's radius is irrelevant for geographical coordinates):

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} \cos(lon_x) \times \cos(lat_x) \\ \sin(lon_x) \times \cos(lat_x) \\ \sin(lat_x) \end{pmatrix}$$

Furthermore the relative frequency $\tilde{f}_x$ in the local grid square is calculated as follows:

$$\tilde{f}_x = \begin{cases} \dfrac{f_x}{F_x} \text{ if } F_x \geq 20 \\ 0.05 \times f_x \text{ otherwise} \end{cases}$$

where:

$$F_x = \sum_{k \in K} f_k$$

with

$$K = \left\{ k \in X \middle| \begin{matrix} lat_x - 0.5 \times gridheight \leq lat_k < lat_x + 0.5 \times gridheight, \\ lon_x - 0.5 \times gridwidth \leq lon_k < lon_x + 0.5 \times gridwidth \end{matrix} \right\}$$

In other words F_x is the sum of frequencies of taxa within the database inside the grid square harbouring match x at its centre. Sample sizes of less than 20 are downweighted.

The vector of the centre of gravity within the vector space is then given by:

$$\mathbf{g} = \frac{\sum_{x \in M} \tilde{f}_x \times \mathbf{x}}{\left| \sum_{x \in M} \tilde{f}_x \times \mathbf{x} \right|}.$$

Note that vector $\mathbf{g}$ has unit length. The vector of the centre of gravity is transformed to geographical coordinates by:

$$lon_g = \arctan\left(\frac{g_2}{g_1}\right)$$

$$lat_g = \arcsin(g_3)$$

The weighted average standard deviation from the centre of gravity is determined by calculating the great circle distance, using the high-precision Haversine formula (Sinnott 1984), between the centre of gravity and each location of a best match, by squaring each distance to a match and multiplying it by its local frequency, and then by extracting the square root of the sum of squares.

Results

For any given human mtDNA control region sequence, mtradius interpolates, between local allele frequencies available in the database, a centre of gravity which serves as a reference point for a dispersion estimate. In order to confirm whether the centre of gravity captures a biologically meaningful property, we tested whether it might estimate the location of a queried sequence. The accuracy of this estimated location will depend on several factors, including sequencing quality in the database (cf. Röhl et al. 2001), sample coverage in the database, mutational parallelisms and choice of the interpolation procedure. Assuming that the attested maternal ancestry of an individual may be approximated by the location of the centre of gravity of his or her mtDNA type, we tested the reliability of our estimated locations by deleting 100 individuals of known origin (as defined in Röhl et al. 2001) at random from the database and then searching for their mtDNA sequence in the remaining database. The deviation (in kilometres) of the estimated from the real maternal geographic origins of these 100 individuals can be seen on the X-axis of Fig. 1. The average accuracy is 2,027 km and 40% of individuals are located within 0–1,000 km of their actual origins and 66% within 0–2,000 km. These results indicate that the centres of gravity generally yield plausible locations for the maximal frequencies of specific alleles, at least on the continental scale.

The current geographic spread of an allele depends on the evolutionary age of the allele, the migration rate and on the mutation rate. In order to distinguish widespread mtDNA types, where the assumption of a continental spread might be sufficiently accurate, from locally spread mtDNA types, for which a local database needs to be consulted before calculating matching frequencies, the mtradius programme offers a standard deviation for each centre of gravity. The standard deviation for the worldwide sample of 100 individuals is given on the Y-axis of Fig. 1 and as expected, the least-squares regression correlates positively (r = 0.48) with the empirically determined deviation (X-axis). Note the concentration of genotypes that have a dispersion of about 1,500 km around their centre of gravity. Many of them represent founder types for whole continents, such as the Cambridge reference sequence, other H sequences and the basic K sequence which are prehistoric founder types for Europe (Richards et al. 1996). A threshold of e.g. 600 km of standard deviation distinguishes these widespread sequences from more local ones below in the plot of Fig. 1. Inevitably, some widespread sequences slip through the 600 km-dispersion "filter" due to inadequate sampling or due to human migration, namely those sequences empirically found to be up to several thousand kilometres distant from the estimated centre of gravity; these erratics are entered in the bottom right rectangle of the diagram below the dotted 600 km line. From the forensic point of view, these potential misclassifications would result in the choice of a local database rather than a continental one, but this choice would

Table 1 mtDNA sequences found in western Sicily and in Greece (*sca* Sciacca, *agr* province of Agrigento, *eub* Euboea, *kio* Chios, *ecr* east

	16037	16041	16051	16064	16066	16067	16091	16071	16086	16092	16093	16104	16111	16114	16129	16134	16136	16140	16145	16147	16148	16153	16163	16167	16168	16169	16171	16172	16174	16176	16178	16179	16181	16182	16183	16184	16186	16187	16189	16192	16193	16195	16204	16209	16211	16212	16217	16223	16224	16229	16231	16232
CRS	A	A	A	T	A	C	C	C	T	T	T	C	C	C	T	G	C	T	T	G	C	C	G	A	C	C	C	A	T	C	C	T	C	A	A	A	C	C	C	T	C	C	T	T	T	C	C	T	T	T	C
sca 40	G	.																																																				
sca 51		G																																			T																
sca 88		G																																		T																
sca 13			G																															C			C														
eub 3			G																																																
kio 47			G																																																
agr 137	N	N		C							C																T																								
sca 10					G																																														
sca 33					G																																														
sca 58					G																																														
sca 79					G																																														
arg 46						T																												C			C														
eub 60						T																														T															
ecr 26							T			C										A																															
arg 41							T								C			C																																	
agr 3442	N	N					T								C					A																															
agr 130	N	N					T								C					A																															
agr 214	N	N					T								C					A										C																			
agr 3492	N	N					T								C					A										C																			
agr 121	N	N					T								C																																			
sca 31							T								C																																			
agr 212	N	N					T								C																																			
emc 42							T								C																																			
agr 209	N	N					T								C																																			
sca 4							T								C																																			
sca 60							T								C																																			
sca 81							T								C																																			
sca 87							T								C																																			
eub 41							T								C																																			
sca 57							T								C																																			
agr 104	N	N					T								C																																			
agr 3576	N	N					T								C																																			
eub 16							T								C																																			
agr 4306	N	N					T								C																																			
emc 21							T								C																																			
eub 30							T								C																																			
ecr 16								T																																											
sca 23									C																												T														
sca 27									C																												T														
lak 17										C																																								
lak 9										C																																								
agr 127	N	N									C					A																																			
kio 21											C					A																																		
arg 55											C																										C														
emc 24											C																																								
agr 106	N	N									C																																							
agr 113	N	N									C																																							
agr 163	N	N									C																																							
agr 177	N	N									C																																							
agr 187	N	N									C																																							
agr 189	N	N									C																																							
agr 220	N	N									C																																							
agr 3269	N	N									C																																							
emc 45											C																																							
sca 7											C																																							
sca 64											C																																							
sca 66											C																																							
sca 72												T																								C	C														
eub 32													T																																					
agr 3411	N	N												T																																				
agr 3412	N	N												T																																				
sca 37															C	A																			C	C														A
arg 42															C			C																																
arg 53															C			C																																
eub 18															C			C																																
ecr 18															C				T																			T												
lak 2															C						A																												
agr 3395	N	N													C						A																												
agr 196	N	N													C							G																											
kio 12															C							G																											
sca 48															C							G																											
sca 55															C							G																											
sca 61															C							G																											
sca 62															C							G																											
agr 183	N	N													C								T																											
emc 43															C														C																				
sca 34															C																																			
agr 203	N	N													C																																			
agr 204	N	N													C																																			
kio 1															C																																		
sca 69															C																																		
arg 52															C																																		
agr 201	N	N													C																																			
emc 22															C																																		
kio 46															C																																		
sca 2															C																																		
sca 73															C																																		
sca 82															C																																		
lak 13															C																																		
eub 17																A	T																																
eub 54																A				A																													
sca 8																A													C																			
emc 23																A																			C	C													
sca 26																A																			C	C													
agr 186	N	N														A																																		
sca 12																A																																	
sca 21																A																																	
sca 38																A																																	
sca 85																A																																	
agr 215	N	N														A																																		
agr 115	N	N														A																																		
lak 15																	T																																	
agr 99	N	N																		A															G															
agr 107	N	N																		A															G															
agr 198	N	N																		A															G															
ecr 29																				A															G															
kio 45																				A																														
sca 3																			T																															
sca 44																			T																															
agr 208	N	N																			T																													
kio 43																						N		T				C																				
sca 15																							T				G																				
sca 49																							T				G																				
kio 42																								T																								
lak 6																								T																								
kio 22																									T																							
kio 44																									T																							
lak 10																										T																						
ecr 19																											C		C	C				C														
ecr 28																												T	G	C				C														
sca 63																												T																				
lak 16																												T																				

Crete, *arg* Argolis, *emc* east Macedonia, *lak* Lakonia)

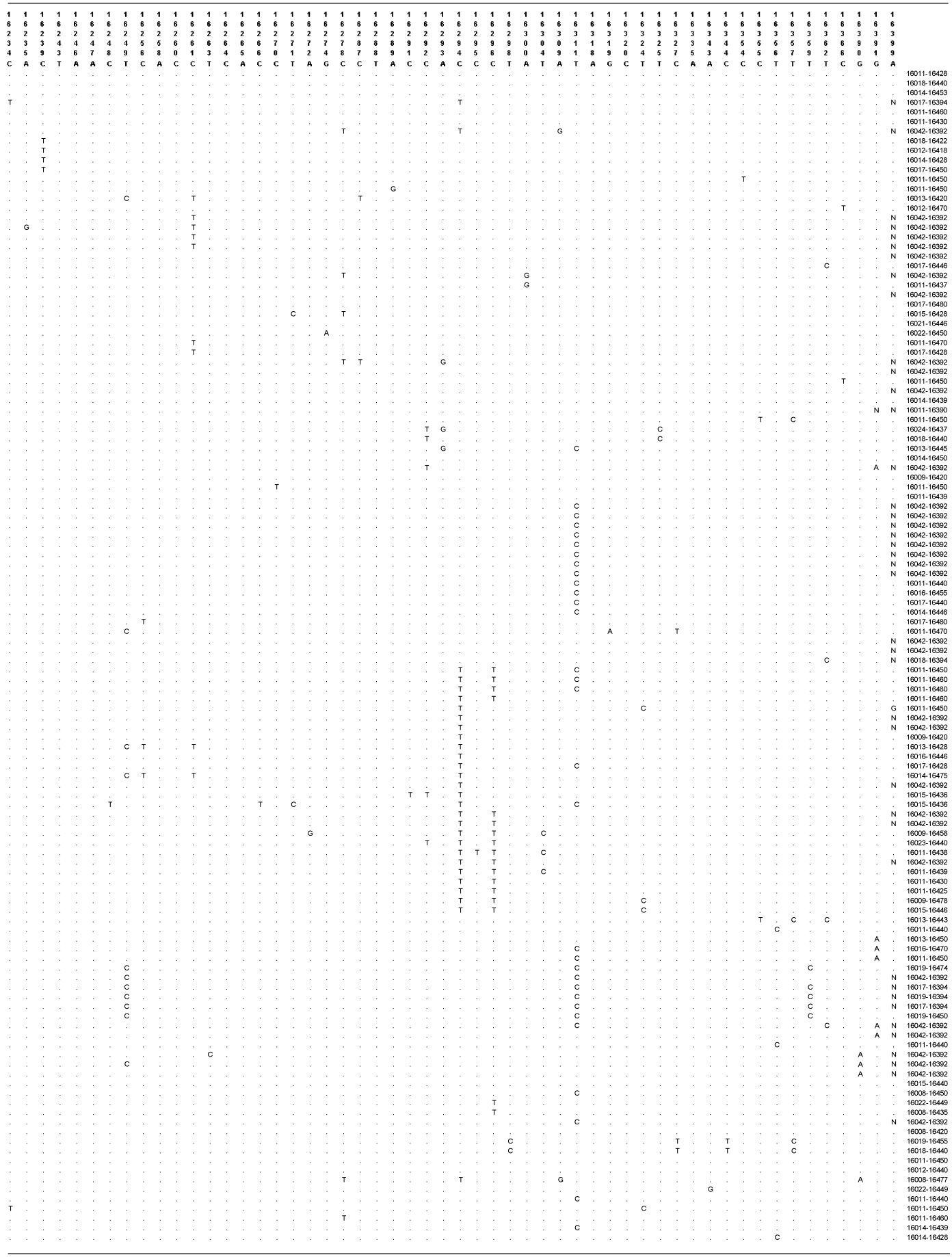
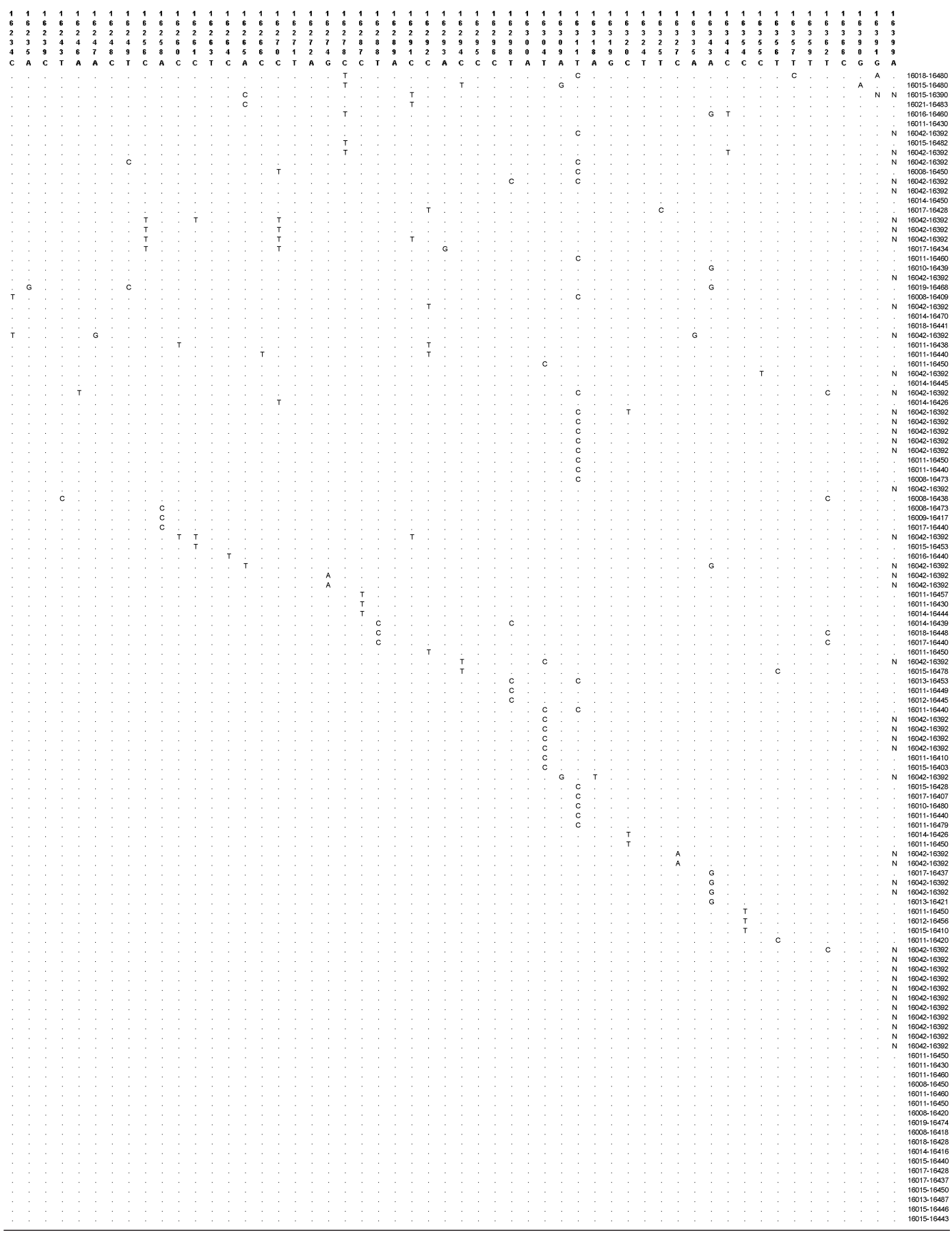


Table 1 (continued)

	CRS	160371A	160411A	160511A	160644T	160666A	160677C	160691C	160711C	160816T	160902T	160933T	161041C	161111C	161146C	161216T	161299G	161341C	161366T	161400T	161455G	161477C	161488C	161533G	161663A	161677C	161688C	161699C	161711A	161721T	161741C	161767C	161788T	161799C	161811A	161822A	161833A	161844C	161866C	161877C	161899T	161922C	161933C	161955T	162009T	162144C	162217T	162218C	162221C	162222C	162223C	162244T	162299T	162311T	162320C
sca	14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.
sca	74	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	C	.	.	.	.	.	.	.	.	T	.	.	.	.
sca	39	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	86	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
arg	44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
emc	29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	194	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	77	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	C	.	.	.	.	T	.	.	.	.
agr	154	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
agr	213	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
kio	31	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	110	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	94	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
lak	8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	32	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	T	.	.	.	.	.
agr	160	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	95	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	157	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
eub	22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	78	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	205	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	18	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	C	.
sca	20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.
agr	168	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	T	C	.	.	.
sca	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
sca	24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.
agr	103	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	C	.	.	.	.
lak	11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
eub	0	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
emc	20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
agr	184	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
sca	83	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.
agr	195	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
sca	65	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
agr	162	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
agr	3394	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
agr	199	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
agr	210	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
agr	211	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
arg	43	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
emc	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
sca	25	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.
agr	125	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.
arg	49	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
kio	2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
kio	11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
kio	41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	185	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	90	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	35	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	191	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	155	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	156	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
arg	50	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
arg	51	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
lak	3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	42	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	70	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
eub	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	119	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	28	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	89	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
eub	7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
lak	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ecr	24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	101	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	111	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	192	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	3404	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
emc	44	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	84	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	149	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
arg	47	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
emc	40	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
emc	41	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
lak	12	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	80	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
arg	45	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ecr	15	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	167	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	170	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	43	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	105	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	159	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ecr	27	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
kio	32	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	11	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	50	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
lak	5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	148	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	108	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	118	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	150	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	164	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	193	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	197	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	219	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	221	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	226	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
agr	58	N	N	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
arg	48	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
arg	54	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
ecr	21	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
eub	4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
eub	24	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
lak	7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	19	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	29	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	36	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	45	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	47	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	54	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	75	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	91	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
sca	92	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.


16234C 16235A 16239C 16243T 16246A 16247A 16248C 16249T 16256C 16258A 16260C 16261C 16263T 16264C 16265A 16266C 16270C 16271T 16272A 16274G 16278C 16287C 16288T 16289A 16291C 16292C 16293A 16294C 16295C 16296C 16298T 16300A 16304T 16309A 16311T 16318A 16319G 16320C 16324T 16325T 16327C 16335A 16343A 16344C 16354C 16355C 16356T 16357T 16359T 16362T 16366C 16380G 16391G 16399A
16018-16480
16015-16480
16015-16390
16021-16483
16016-16460
16011-16430
16042-16392
16015-16482
16042-16392
16042-16392
16008-16450
16042-16392
16042-16392
16014-16450
16017-16428
16042-16392
16042-16392
16042-16392
16017-16434
16011-16460
16010-16439
16042-16392
16019-16468
16008-16409
16042-16392
16014-16470
16018-16441
16042-16392
16011-16438
16011-16440
16011-16450
16042-16392
16014-16445
16042-16392
16014-16426
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16011-16450
16011-16440
16008-16473
16042-16392
16008-16436
16008-16473
16009-16417
16017-16440
16042-16392
16015-16453
16016-16440
16042-16392
16042-16392
16042-16392
16011-16457
16011-16430
16014-16444
16014-16439
16018-16448
16017-16440
16011-16450
16042-16392
16015-16478
16013-16453
16011-16449
16012-16445
16011-16440
16042-16392
16042-16392
16042-16392
16042-16392
16011-16410
16015-16403
16042-16392
16015-16428
16017-16407
16010-16480
16011-16440
16011-16479
16014-16426
16011-16450
16042-16392
16042-16392
16017-16437
16042-16392
16042-16392
16042-16392
16013-16421
16011-16450
16012-16456
16015-16410
16011-16420
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16042-16392
16011-16450
16011-16430
16011-16460
16008-16450
16011-16460
16011-16450
16008-16420
16019-16474
16008-16418
16018-16428
16014-16416
16015-16440
16017-16428
16017-16437
16015-16450
16013-16487
16015-16446
16015-16443

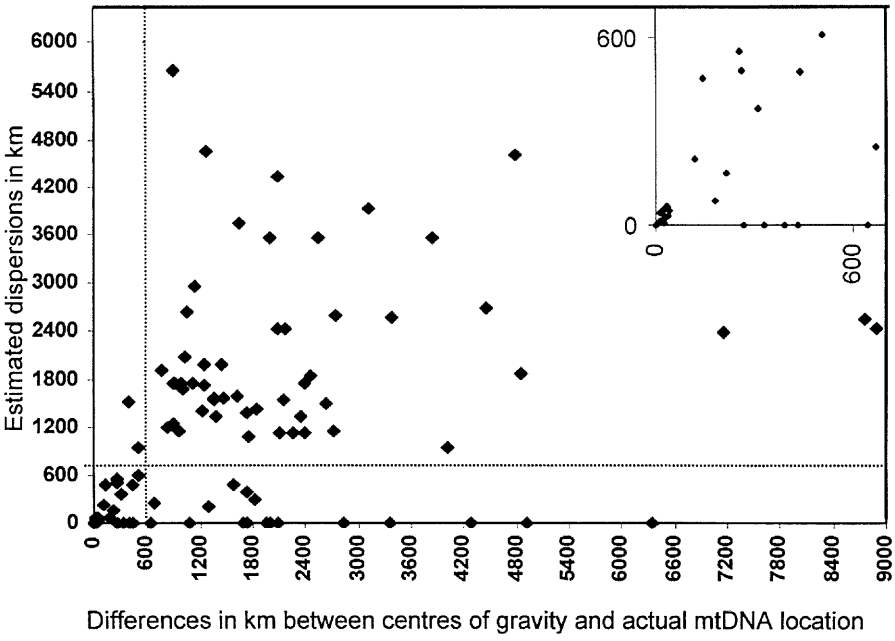


Fig. 1 Estimated dispersions (Y-axis) and differences between actual and estimated locations (X-axis) for each of 100 mtDNA types randomly drawn from the worldwide mtDNA database. The dispersion is expressed as a standard deviation around the centre of gravity of the genetically closest mtDNA matches (see Methods) after elimination from the database of the randomly drawn mtDNA types

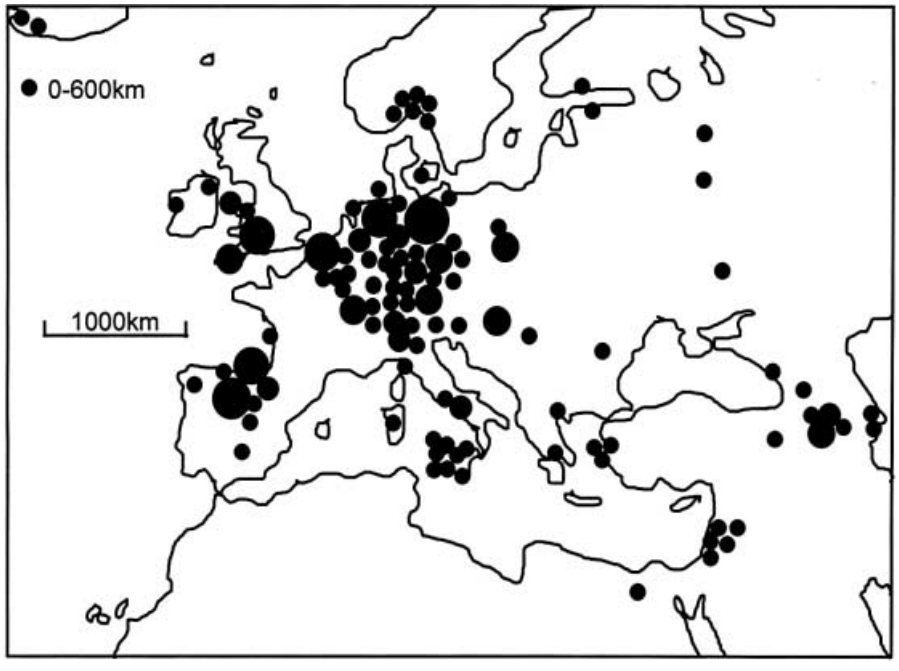


Fig. 2 Global centres of gravity for 1,123 mtDNA sequences from one north German village. Each circle represents a centre of gravity for a sequence type found in the German village, with the circle area proportional to the number of villagers with that sequence. Only those centres of gravity with estimated dispersions of less than 600 km are shown. Twelve villagers have their centres of gravity in Asia, and two villagers have their centres of gravity in Africa and thus do not appear on the map

typically not work against the suspect if the sequence indeed were widespread; the choice would operate in favour of the suspect if the misclassification were due to an exceptional migration event or parallel mutation event. An ambiguous search result is obtained when exact matches are not found in the database and the next best matches are geographically widespread: whether or not the queried mtDNA type is localised cannot then be answered without enlarging the database.

The validity of our distinction between local and widespread types can be demonstrated by geographically analysing the mtDNA of 1,123 males with German family names sampled from a single village in north Germany, near Brunswik (Pfeiffer et al. 2001). In the first step, we deleted the 1,123 German villagers from the database and then queried each villager one by one. Of the total of 1,123 villagers, 174 have mtDNA types whose dispersion is estimated by mtradius to be less than 600 km around the estimated centre of gravity. When these 174 "low-dispersion" centres of gravity are plotted on a map (Fig. 2), it is reassuring to see that 68 (6.1%) of individuals cluster their low-dispersion centres of gravity in and around Germany whereas 106 individuals have their low-dispersion centres of gravity elsewhere – proportions suggested above in Fig. 1.

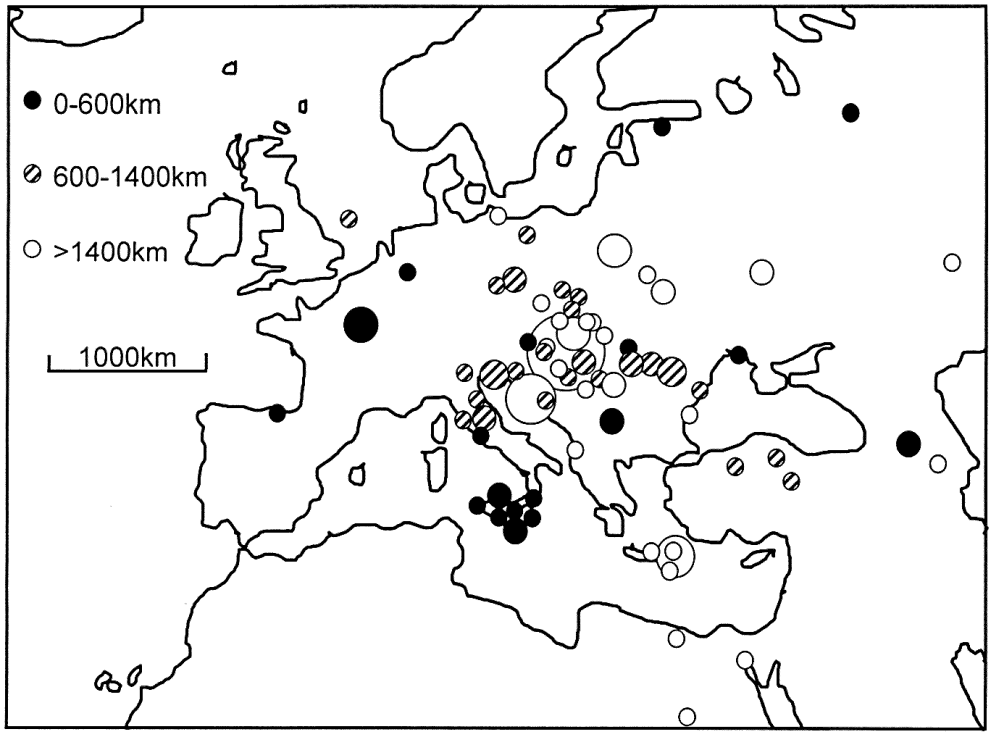


Fig. 3 Global centres of gravity for 159 mtDNA sequences from the Province of Agrigento (west Sicily) including the town of Sciacca. Each circle represents a centre of gravity for a sequence type found in the Agrigentini, with the circle area proportional to the number of Agrigentini with that sequence. Eight centres are located in Asia and three in Africa and therefore do not appear on this map. The circle shading indicates the standard deviation in kilometres of a centre of gravity as calculated by mtradius

Incidentally, the overall centre of gravity for the major cluster of low-dispersion centres of gravity in Fig. 2 is a point near Frankfurt, only 280 km distant from the village. In other words it appears possible to locate an unknown population sample (even a modern mixed one as in this case) fairly accurately on the basis of their mtDNA types, which will be of interest for population genetic applications. In the German example, there are also minor clusters around Kurdistan, south Italy, Spain and Britain, which contain an average of 13 individuals per cluster (1.2% of the total sample, compared to 6.1% for the German cluster). Even if we take the extreme view that all these minor clusters are artefacts of the database coverage and of the interpolation procedure rather than real evidence of recent migration, then the signal-to-noise ratio of the cluster criterion for identifying sample origins would be 5:1 in this case, which appears quite useful considering the very short length (270 bp) of HVR1 sequences.

It may be argued that the clustering of centres of gravity in Germany might be an artefact of the relatively good sample coverage in and around Germany in the database (of the 4,006 active sequences between Iceland and Turkey, nearly 900 are from Germany, Austria, Switzerland and north Italy). To explore this possibility, we analysed in the same way 159 west Sicilians (from Sciacca and the surrounding Province of Agrigento) as a negative control. As can be seen in Fig. 3, only one Sicilian has a centre of gravity in Germany, while the rest cluster in Sicily and elsewhere. Uneven sample coverage therefore does not appear to be a major confounding factor. In the Sicilian analysis (Fig. 3) we have also included those centres of gravity whose dispersion is estimated to be greater than 600 km. Most (93%) of these cluster in or around Europe, clearly identifying the west Sicilians as European, while a few centres of gravity are found to be typical African (3 Sicilians, i.e. 2%) or Asian (8 Sicilians, i.e. 5%) sequences even though the Sicilians were traced maternally to the Province of Agrigento for two or three generations. This underlines that phenotype is not an infallible guide to a reference population harbouring maximal allele frequency.

In some cases, a centre of gravity interpolated into an unsampled area can be a useful indication for a hitherto unsampled frequency maximum. However, it should be understood that a standard deviation associated with a centre of gravity will not represent elongated distributions of mtDNA matches very well; the interpolation should therefore not be used without first consulting the geographic spread of relative frequencies, included as an option in the database.

Conclusions

Using the presently available mtDNA database, it is straightforward to assign mtDNA types to their continent of origin in the great majority of cases (average accuracy of about 2,000 km) and many of these sequences can be shown to be widespread within their respective continents, allowing the choice of a continental database for maximum chance matching probabilities when a suspect's mtDNA matches that of a crime stain. There is however a residue (approximately 5–10% in European HVR1 sequences) of locally dispersed mtDNA types, in which

case a local population database should be chosen when calculating chance matching in order to give the suspect the maximum benefit of doubt.

Acknowledgements We thank Bob Chamberlain (Jet Propulsion Laboratory, NASA) for guidance with the spherical trigonometric formulae, Martin Richards (Department of Chemical and Biological Sciences, University of Huddersfield) and Olga Rickards (Department of Biology, University of Rome "Tor Vergata") for information on east European and Sicilian data, respectively, and Vincent Macaulay (Department of Statistics, University of Oxford) and two anonymous reviewers for valuable comments.

References

Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465

Calì F, Le Roux MG, D'Anna R, Flugy A, De Leo G, Chiavetta V, Ayala GF, Romano V (2001) MtDNA control region and RFLP data for Sicily and France. Int J Legal Med 114:229–231

Cavalli-Sforza LL, Minch E (1997) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 61:247–251

Helgason A, Hickey E, Goodacre S, Bosnes V, Stefánsson K, Ward R, Sykes B (2001) MtDNA and the islands of the north Atlantic: estimating the proportions of Norse and Gaelic ancestry. Am J Hum Genet 68:723–737

Lewontin RC (1972) The apportionment of human diversity. Evol Biol 6:381–398

Melton T, Clifford S, Kayser M, Nasidze I, Batzer M, Stoneking M (2001) Diversity and heterogeneity in mitochondrial DNA of North American populations. J Forensic Sci 46:46–52

National Research Council Committee (1992) DNA technology in forensic science. National Academy Press, Washington DC

National Research Council Committee (1996) The evaluation of forensic DNA evidence. National Academy Press, Washington DC

Nature Genetics Editorial (2000) Census, race and science. Nat Genet 24:97–98

Pfeiffer H, Forster P, Ortmann C, Brinkmann B (2001) The results of an mtDNA study of 1200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework. Int J Legal Med 114:169–172

Pult I, Sajantila A, Simanainen J, Georgiev O, Schaffner W, Pääbo S (1994) Mitochondrial DNA sequences from Switzerland reveal striking homogeneity of European populations. Biol Chem Hoppe Seyler 375:837–840

Richards M, Côrte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S, Hedges R, Bandelt H-J, Sykes B (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. Am J Hum Genet 59:185–203

Röhl A, Brinkmann B, Forster L, Forster P (2001) An annotated mtDNA database. Int J Legal Med 115:29–39

Simoni L, Calafell F, Pettener D, Bertranpetit J, Barbujani G (2000) Geographic patterns of mtDNA diversity in Europe. Am J Hum Genet 66:262–278

Sinnott RW (1984) The virtues of the Haversine. Sky Telescope 68:159

Torroni A, Bandelt H-J, D'Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M-L, Bonné-Tamir B, Scozzari R (1998) MtDNA analysis reveals a major late Palaeolithic population expansion from southwestern to northeastern Europe. Am J Hum Genet 62:1137–1152

Weber M (1922) Wirtschaft und Gesellschaft. Grundriß der Sozialökonomik III. Verlag JCB Mohr (Paul Siebeck), Tübingen, pp 216–222

Wittig H, Augustin C, Baasner A, Bulnheim U, Dimo-Simonin N, Edelmann J, Hering S, Jung S, Lutz S, Michael M, Parson W, Poetsch M, Schneider PM, Weichhold G, Krause D (2000) Mitochondrial DNA in the central European population. Human identification with the help of the forensic mt-DNA D-loop-base database. Forensic Sci Int 113:113–118