Assigning African elephant DNA to geographic region of origin: Applications to the ivory trade

Samuel K. Wasser*[†], Andrew M. Shedlock*[‡], Kenine Comstock*[§], Elaine A. Ostrander[§], Benezeth Mutayoba[¶], and Matthew Stephens[‖]

*Department of Biology, Center for Conservation Biology, University of Washington, Box 351800, Seattle, WA 98195; [§]Clinical Research and Human Biology Divisions, Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, D4–100, Seattle, WA 98109-1024; [¶]Department of Veterinary Physiology, Biochemistry, Pharmacology, and Toxicology, Sokoine University of Agriculture, P.O. Box 3017, Morogoro, Tanzania; and [‖]Department of Statistics, University of Washington, Seattle, WA 98195

Resurgence of illicit trade in African elephant ivory is placing the elephant at renewed risk. Regulation of this trade could be vastly improved by the ability to verify the geographic origin of tusks. We address this need by developing a combined genetic and statistical method to determine the origin of poached ivory. Our statistical approach exploits a smoothing method to estimate geographic-specific allele frequencies over the entire African elephants' range for 16 microsatellite loci, using 315 tissue and 84 scat samples from forest (*Loxodonta africana cyclotis*) and savannah (*Loxodonta africana africana*) elephants at 28 locations. These geographic-specific allele frequency estimates are used to infer the geographic origin of DNA samples, such as could be obtained from tusks of unknown origin. We demonstrate that our method alleviates several problems associated with standard assignment methods in this context, and the absolute accuracy of our method is high. Continent-wide, 50% of samples were located within 500 km, and 80% within 932 km of their actual place of origin. Accuracy varied by region (median accuracies: West Africa, 135 km; Central Savannah, 286 km; Central Forest, 411 km; South, 535 km; and East, 697 km). In some cases, allele frequencies vary considerably over small geographic regions, making much finer discriminations possible and suggesting that resolution could be further improved by collection of samples from locations not represented in our study.

Trade in wildlife products such as ivory, bush meat, and whale meat is capable of rapidly decimating species on a continent-wide scale (1–3). Consequences of poaching often go undetected until the damage becomes too severe to correct, especially for wildlife in difficult to observe habitats such as rain forests or open ocean. The ability to determine origin of wildlife products could help address these problems, providing early warning of where greater law enforcement is needed, and helping guide international decisions regarding delisting species and resanctioning their legal trade. We describe a combination of genetic, sampling, and statistical methods for inferring the geographic origin of elephant DNA that can greatly assist such management problems.

High demand for ivory reduced the African elephant (*Loxodonta africana*) population from 1.3 million to 600,000 individuals between 1979 and 1987 (4). This circumstance prompted the Convention on International Trade in Endangered Species (CITES) to implement a ban on ivory trade in 1989. However, international pressure persisted to resanction trade even though the illicit ivory market continued to thrive (5–7). Three of the largest ivory seizures since the trade ban occurred since June 2002 (Singapore, Hong Kong, and Shanghai, China). Elephant poaching in Central Africa forests has been particularly severe in recent years, where poor visibility has hindered monitoring. Much of this ivory is presumed to be smuggled into international markets by West African countries (8). Monitoring the origin of ivory passing through the major ivory markets around the world would greatly assist efforts to contain such trade.

Our methods rely on (*i*) noninvasive techniques to acquire DNA from scat (9–10), enabling rapid development of a continent-wide geographic map of elephant allele frequencies; (*ii*) the ability to acquire DNA from small amounts of ivory taken anywhere along the length of the tusk (11); and (*iii*) spatial smoothing methods to estimate allele frequencies at any location in a continuous region, using genotypes of reference samples from a number of sampling locations. Our smoothing approach allows us to improve on standard assignment methods (SAMs) (12), both in terms of assignment accuracy, and more importantly, allowing assignment of samples to locations where no reference samples are available. The latter is especially important in our application, because ivory seizures may have come from such locations.

Our method accurately discriminates among DNA samples originating from forest (*Loxodonta africana cyclotis*) and savannah elephants from the four major regions of Africa (West, Central, South, and East). In some cases, allele frequencies vary considerably over small geographic regions, making much finer distinctions possible, and suggesting that resolution could be further improved by collection of samples from locations not represented in our study. We also present the first compelling, to our knowledge, genetic evidence for continued hybridization between forest and savannah elephants, although such hybridization appears to be rare in our samples.

Materials and Methods

Samples. Mitochondrial and microsatellite DNA can be isolated and amplified from small amounts of African elephant ivory taken nearly anywhere along a tusk (11). No attached tissue is required and old samples stored at ambient temperatures can be successfully analyzed. We collected genotypes at 16 microsatellite loci (13–15) by using DNA isolated from 350 tissue samples and 242 fecal samples collected from 28 locations in 14 African countries (Fig. 1). Of these samples, 399 had alleles amplify at seven or more loci, and were used in subsequent analyses. Approximately 85% of Africa's elephants reside in the 14 countries represented in our sample (15).

We used skin biopsy samples analyzed in Comstock *et al.* (14), which were collected by using the method of Karesh *et al.* (16) in full compliance with CITES regulations. Fecal samples were collected with no two samples being closer than 1 km apart to reduce chances of sampling multiple individuals from the same family group. Fecal samples were either preserved in 90% ethanol or 20% DMSO in TNE buffer, recorded for date and location, and transported to the United States in compliance with U.S. Department of Agriculture/Animal and Plant Health Inspection Service regulations.
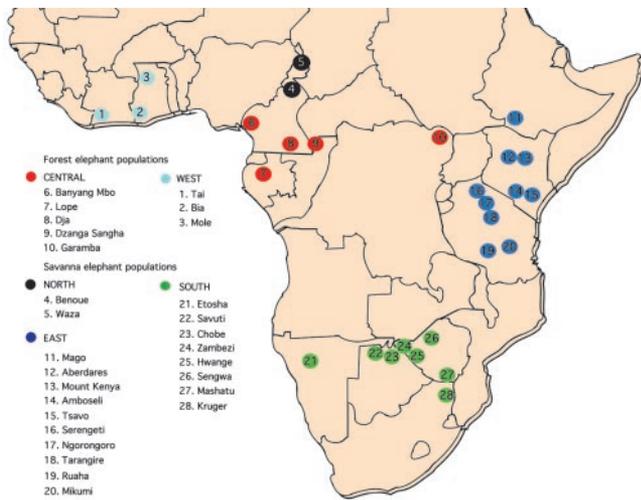
GENETICS

Fig. 1. Map of Africa showing the collection sites divided into five regions: West Africa (cyan), Central forest (red), and Central (black), South (green), and East (blue) savannah.

DNA Extraction and Amplification. DNA was extracted in duplicate from fecal samples by using a QIAquick Stool DNA purification kit (Qiagen, Valencia, CA) followed by a Gene Clean III (Bio101) nucleic acid isolation kit for each template isolation. Negative controls were included for every set of 10 extractions completed. We used standard three-step PCR following an adaptation of the multiple tubes approach (17) to identify/minimize allelic dropout (10, 18). All heterozygotes were scored twice, and homozygous individuals were examined by several independent amplification products from two independent DNA isolations derived from the same fecal sample. This approach yielded consistent consensus scores for all multiple amplifications included in the final data set.

Amplification of target loci from <50 ng of genomic template was achieved by 35 cycles of standard three-step PCR in 25-μl reactions employing 5′-6-carboxy fluorescein-labeled forward oligonucleotides (Applied Biosystems) and *Taq* polymerase antibody hotstart chemistry (Clontech) in an ABI 9700 Gene Amp PCR System. One microliter of PCR products was subjected to fragment analysis by using the GeneScan mode on an ABI capillary array genetic analyzer, model 3100. Allele sizes were scored by using the program GENOTYPER V. 3.7, with categories defined by the weighted average of histograms for each allele size bin (tolerance 0.5 bp).

Raw genotype data from acrylamide vs. Applied Biosystems capillary platforms would be expected to exhibit small allele size differences, consistent across samples within each locus. We reconciled data across platforms in two ways: (*i*) For each locus, allele size differences (acrylamide minus Applied Biosystems capillary platforms) were computed for control DNAs typed on both platforms, to establish a correction factor to be added to each Applied Biosystems call; and (*ii*) empirical population allele frequencies for samples from Benoue and Dzanga Sanga were compared across platforms, revealing a single discrepancy after correction as in (*i*) above: at locus FH40, allele size 229 was frequently observed in acrylamide genotypes, whereas an allele of size 227 was frequently observed in Applied Biosystems genotypes. Because no alleles of similar size occurred on either platform, we concluded these were the same allele.

Statistical Procedures. *Hybrid and assignment analyses.* We used a variation of a standard assignment procedure to assign population of origin to samples, and to identify putative forest–savannah hybrids (see also ref. 19). Let g denote the sample genotype, with alleles at locus l being (g_{l1}, g_{l2}), and let f_{jlk} denote the frequency of allele j at locus l in population k. (For the hybrid analysis, we pool

our sampling locations into five populations: West and Central Forest, and Central, East, and South savannah, according to geographic location and habitat. For assignment analyses, we treat each sampling location as a separate population.) Assuming Hardy–Weinberg equilibrium and linkage equilibrium within each population, the probability of observing this sample, if the two parents came from populations k_1 and k_2, can be written as:

$$L(k_1, k_2; g, f) = \Pi_l p_1(g_{l1}, g_{l2}|k_1, k_2, f),\qquad [1]$$

where

$$p_l(i, j|k_1, k_2, f) = (1 - \gamma)\,[p_l(i|k_1, f)\,p_l(j|k_2, f]$$
$$+ p_l(j|k_1, f)\,p_l(i|k_2, f)$$
$$\text{if } i \neq j, \text{ and}$$
$$p_l(i, i|k_1, k_2, f) = (1 - \gamma)\,p_l(i|k_1, f)\,p_l(i|k_2, f)$$
$$+ 0.5\gamma\,[p_l(i|k_1, f) + p_l(i|k_2, f)],$$

with

$$p_l(i|k, f) = (1 - \delta)f_{jlk} + \delta/m_l.$$

Here m_1 is the number of observed alleles (across all populations) at locus l, δ is a genotyping error probability, and γ is the probability of only one allele amplifying. We fixed $\delta = 0.05$, $\gamma = 0.1$. Although in this case, we obtained qualitatively similar results (data not shown) when we ignored the possibility of genotyping error (i.e., $\delta = 0.0$, $\gamma = 0.0$), in general, allowing for the possibility of genotyping error is an important and often overlooked aspect of these kinds of analyses.

We considered an individual a potential hybrid if the maximum of $L(k_1, k_2)$ over k_1 and k_2 occurred for k_1 in a savannah region and k_2 in a forest region. For standard assignment analyses, we assigned each individual to the population k that maximized $L(k, k)$. In each case, we estimated f_{jlk} as $(n_{jlk} + 1)/(n_{+l+} + m_1)$, where n_{jlk} is the number of times allele j is observed at locus l in population k, and $n_{+l+} = \Sigma_{jk}\, n_{jlk}$. Our smoothing-based assignment procedure also assigns each individual to the population k that maximizes $L(k, k)$ but with f estimated by using the smoothing-based method described below. When analyzing a particular sample, the alleles for that sample were ignored in computation of n_{jlk} (i.e., we used leave-one-out crossvalidation.)

Spatial-smoothing-based estimates of allele frequencies. Our smoothing-based method for estimating f_{jlk} estimates allele frequencies at any location, by using all reference samples, with samples from nearby sampling locations being given more weight. Our method is based on writing

$$f_{jlk}(\theta) = \exp(\theta_{jlk}) \Big/ \sum_{j'} \exp(\theta_{j'lk}),\qquad [2]$$

where the θ values corresponding to each locus-allele combination are assumed to be independent Gaussian processes. More specifically, the θ_{jlk} have a joint normal distribution, with $E(\theta_{jlk}) = \mu_{jl}$, and $Cov(\theta_{jlk}, \theta_{jlk'}) = \sigma_{kk'}$, with θ_{jlk} and $\theta_{j'l'k}$ being independent if $l \neq l'$ or $j \neq j'$. In this model, μ controls the mean allele frequencies (across the study region) at each locus, and the population-specific allele frequencies are allowed to vary about this mean in a spatially correlated way. The value of $\sigma_{kk'}$ controls the (expected) degree of similarity between allele frequencies in populations k and k'. We assume that $\sigma_{kk'}$ depends only on the distance (in kilometers), $d_{kk'}$, between populations k and k', with $\sigma(d; \alpha) = (1/\alpha_0)\exp[-(\alpha_1 d)^{\alpha_2}]$, where $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ are parameters to be estimated. See ref. 20 for other possible parameterizations. The value of α_0 controls the variability at $d = 0$ (i.e., variability of regional frequencies from the mean); $\alpha_0 =$

∞ gives no variation in allele frequencies. The parameters α_1 and α_2 control how the correlations decay with distance. Informally, α_1 controls the natural scale (e.g., kilometers, or tens of kilometers) on which the allele frequencies vary, and α_2 controls how quickly the correlations drop to 0. The case $\alpha_2 = 0$ corresponds to no spatial correlation in frequencies.

The values of α are estimated from the data in a Bayesian way (20). We use independent priors on components of α: $\alpha_0 \sim \Gamma(0.001, 0.001)$; $\log_{10}(\alpha_1) \sim \text{Uniform}(0, 4)$; $\alpha_2 \sim \text{Uniform}[0.1, 2]$. We use independent $N(0, 1/\beta)$ priors for components of μ, and a $\Gamma(0.001, 0.001)$ hyperprior on β. We use Markov Chain Monte Carlo (see *Supporting Text*, which is published as supporting information on the PNAS web site) to sample from the posterior distribution of f, and estimate f_{jlk} by the mean of sampled values. Previous work on modeling spatial variation in allele frequencies includes ref. 21. Our model differs from theirs in several ways, including (*i*) our parameterization (Eq. **1**) treats every allele symmetrically, whereas they choose one allele against which every other allele's frequency is compared; and (*ii*) they use a regression to model θ, whereas we use a Gaussian process.

Continuous assignment method (CAM). Our smoothing-based method can estimate allele frequencies at any location, including locations where no samples are available. We exploit this to develop our CAM, which allows samples to be assigned to any location, and not only locations with reference samples, thus overcoming a limitation of standard assignment tests (see also ref. 22). Let W denote the unknown position of origin of a sample to be located. We estimate W in a Bayesian framework, which accounts for uncertainty in estimated allele frequencies. We placed a uniform prior on W over all parts of the continent inhabited by forest elephants, or by savannah elephants, depending on whether the sample being analyzed was from a forest or savannah region. We extended the Markov Chain Monte Carlo algorithm to include W as an unknown quantity, and obtain a sample from the posterior distribution of W (*Supporting Text*). The median of the latitude and longitude of the sample give an estimate for W, and the spread gives an indication of the precision.

All assignment results were obtained by using leave-one-out crossvalidation, in which each sample in turn was treated as the sample whose location was unknown, whereas the other samples were assumed to have known location.

Results

Previous genetic studies suggest that African elephants may be subdivided into at least two species: forest and savannah (14, 23, 24). However, observational studies have reported hybridization between the two groups on the edges of their range (25). If common, this would impact the method development for tracking ivory samples. To estimate the extent of such hybridization, we computed for each sample a likelihood ratio (LR) to assess if the sample could be the offspring of one forest and one savannah elephant. Only two of our samples, both from Garamba, gave an LR favoring a hybrid origin, suggesting that in our sample hybridization is rare. However, one of these samples had overwhelming support for hybrid (central forest × central savannah) origin (LR >14,000 vs. pure central savannah; LR >4.5 million vs. pure central forest), providing compelling genetic evidence for continued hybridization between the groups. Furthermore, except for these putative hybrid individuals, the forest or savannah origin of every sample was correctly determined. We therefore treated the two groups separately for subsequent analyses.

We used our smoothing assignment procedures to attempt to infer, from the genotype data, the sampling location of each sample (Tables 1 and 2). Forest elephants were accurately assigned to their actual forest of origin (Table 1), suggesting considerable population structure, despite the fact that all West samples, and 34% of the Central forest samples, were from DNA

Table 1. Classification matrix for samples from forest locations by using the smoothing-based assignment method

	Estimated location*								Accuracy, %
	TI	BI	MO	BM	DJ	DS	LO	GA	
TI	2	0	0	0	0	0	0	0	100
BI	0	9	0	0	0	0	0	0	100
MO	2	0	9	1	0	0	0	0	75
BM	0	0	1	10	0	0	0	0	91
DJ	2	0	0	0	6	1	0	0	67
DS	4	0	1	2	1	34	8	4	63
LO	0	0	0	0	1	1	13	0	87
GA	2	0	0	0	0	2	0	16	80

Average accuracy, 83%.

*TI, Tai; BI, Bia; MO, Mole; BM, Banyang Mbo; DJ, Dja; DS, Dzanga Sangha; LO, Lope; GA, Garamba.

isolated from scat, which amplified less reliably than tissue samples (0.22 vs. 0.06 missing loci on average; see Table 3). Fortunately, DNA appears to be well preserved in ivory (11), providing higher amplification success than from scat.

Savannah elephants were more difficult to assign to their specific sampling location. However, when samples were wrongly assigned the estimated location was typically near to the actual location. Further, some locations [e.g., Etosha (ET), Mashatu (MA), and to a lesser extent, Kruger (KR)] had a large proportion of samples correctly assigned, suggesting that they are genetically somewhat distinct from other locations we sampled. ET and MA are separated from other sampling locations by distance and habitat. KR may have had a huge influx of elephants from Mozambique in the 1960s (26).

To demonstrate the advantages of our smoothing method, we compare it with results (Tables 4 and 5, which are published as supporting information on the PNAS web site) from a SAM (see *Materials and Methods*). (Several alternative SAMs exist, and they would give quantitatively different results. However, any method that estimates allele frequencies in each population separately, using only samples from that population, will suffer the problems that we identify here due to small reference sample sizes.) Our smoothing-based method gives higher average assignment accuracy than the SAM (83% vs. 68% for forest locations and 35% vs. 30% for savannah). However, besides producing lower assignment accuracy, the SAM has two other important problems that are alleviated by smoothing. First, it has a strong systematic bias against assigning samples to locations where reference sample sizes are small. For example, in contrast to the smoothing method, the SAM assigned none of the 399 samples to the 4 sampling locations (Tai, Mago, Mount Kenya, and Ruaha) with <5 reference samples. The SAM tends not to assign samples to these locations because small reference samples give inaccurate allele frequencies there, resulting in artifically small likelihoods for observed genotypes; SAM assigns each sample to locations giving the highest likelihood for its genotype. To further test this premise, we compared assignment results for samples from the DS location ($n = 54$), by using (*i*) the full reference sample at DS, and (*ii*) a reduced reference sample of only three individuals. When the reduced reference sample is used, the SAM assigns 0/51 of the remaining samples to DS, in striking contrast to the 47/54 that were assigned there when using the full reference sample (Table 4). In comparison, reducing the size of the reference sample at DS has a much less dramatic effect on the smoothing method (corresponding numbers are 9/51 and 34/54, see Table 1).

The SAM's tendency to assign samples to locations with larger reference samples actually increases the total number of reference samples correctly assigned because, by definition, locations with larger reference samples are overrepresented in the samples being assigned. In our application, what matters in practice is not accuracy

Table 2. Classification matrix for samples from savannah locations by using the smoothing method

	Estimated location*																				Accuracy, %
	BE	WA	MG	AB	MK	AM	TZ	SE	NG	TA	MI	RU	SW	HW	ZZ	CH	SA	MA	KR	ET	
BE	9	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	69
WA	5	15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	71
MG	0	0	2	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	50
AB	0	0	3	1	2	1	4	1	1	0	1	0	3	0	0	0	1	0	0	1	5
MK	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	33
AM	0	0	2	0	1	3	1	1	1	0	0	1	0	0	0	0	1	0	1	2	21
TZ	1	0	4	0	4	1	3	1	0	0	1	0	1	0	0	0	0	0	0	0	18
SE	0	1	3	2	3	0	3	0	1	1	0	0	1	0	0	0	0	2	1	0	0
NG	0	0	1	1	0	0	0	0	2	0	2	4	0	1	0	0	1	0	0	2	13
TA	0	0	1	0	1	1	0	2	1	1	2	2	1	0	0	0	0	0	0	0	8
MI	0	0	0	0	1	0	1	0	0	1	4	1	2	0	0	0	0	1	1	0	33
RU	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	50
SW	0	0	1	0	0	0	0	0	1	0	1	6	2	0	1	0	2	0	0	2	12
HW	0	0	1	0	0	0	0	0	0	0	0	1	1	0	1	1	2	2	1	1	0
ZZ	0	0	0	0	0	0	0	0	0	0	0	0	0	1	8	1	2	0	1	0	61
CH	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2	1	6	1	1	0	8
SA	0	1	0	0	0	0	0	0	0	0	0	0	2	1	3	1	1	0	2	4	6
MA	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	19	0	0	90
KR	0	0	0	0	0	0	0	0	0	0	1	0	0	2	1	0	0	1	11	0	68
ET	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	2	11	68

Average accuracy, 35%.

*BE, Benoue; WA, Waza; MG, Mago; AB, Aberderes; MK, Mount Kenya; AM, Amboseli; TZ, Tsavo; SE, Serengeti; NG, Ngorongoro; TA, Tarangire; MI, Mikumi; RU, Ruaha; SW, Sengwa; HW, Hwange; ZZ, Zambezi, CH, Chobe; SA, Savuti; MA, Mashatu; KR, Kruger; ET, Etosha.

on our reference samples, but accuracy for future samples (e.g., from ivory seizures). Most likely, such samples will tend to come from locations with fewer reference samples because it may be harder to find reference samples in locations where poaching has dramatically reduced the elephant population. For this reason, the total number of reference samples correctly assigned is not a good measure of how the methods will perform on future samples. If we assume that future samples are equally likely to come from any of our sampling locations (which both assignment schemes assume implicitly by assigning samples to the location with the highest likelihood) then the average accuracy across sampling locations, quoted above and in Tables 1–4, estimates the expected accuracy of the methods on future samples.

A second problem with the SAM is that the artificially small likelihoods arising from inaccurate allele frequency estimates can cause gross overconfidence in incorrect conclusions. One measure of the confidence in an assignment to one location vs. another is the LR for the sample genotype in the two locations. If the LR for the assigned population vs. the true population is large (>1,000 for instance), then we might confidently, but wrongly, conclude that the true population is not the source of the sample. This finding is important in our application because it may be helpful to exclude a location as a potential source for a tusk even if the actual source remains uncertain. Of the samples that come from sampling locations with <10 reference samples, 9/29 have an LR >1,000, favoring the wrong location for the SAM. In contrast, 0/29 have an LR >1,000 by using the smoothing method. Similarly, for the experiment with a reduced reference sample in DS, 26/51 samples had LRs >1,000, favoring the wrong location by the SAM, vs. 2/51 by the smoothing method.

Our smoothing assignment method alleviates some of the problems with the SAM by improving allele frequency estimates at locations with small reference samples (and, in fact, at all locations in our study; see *Supporting Text*, and Fig. 4, which is published a

Table 3. Number of kilometers within which 20%, 50%, and 80% of samples could be correctly located by the continuous assignment method

Area	Method*	20%	50%	80%	Tissue/scat[†]	
					Samples	No. of loci
Overall	Included	212	499	932	315/84	15.1/12.5
	Excluded	404	731	1,263		
West	Included	39	135	331	0/23	NA/12.8
	Excluded	392	559	1,891		
Central forest	Included	203	411	720	72/37	15.3/11.0
	Excluded	432	725	1,117		
Central savanna	Included	98	286	405	26/6	14.8/13.8
	Excluded	247	338	450		
East	Included	339	697	1,293	95/18	15.3/14.8
	Excluded	431	844	1,450		
South	Included	227	535	933	120/0	15/NA
	Excluded	459	836	1,379		

NA, not applicable.

*The two methods (included and excluded) refer to all neighboring samples from that subpopulation were included versus excluded from the calculation of geographic-specific allele frequencies used to assign individuals from that location (see text).

[†]Number of tissue and scat samples and mean number of loci successfully amplified per respective sample.
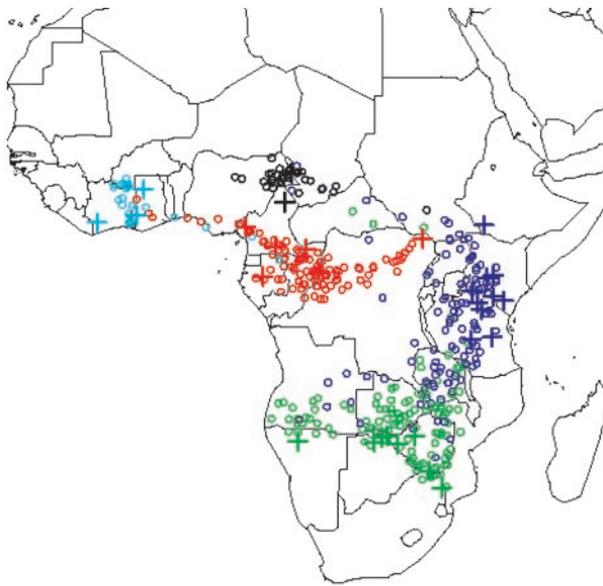
Fig. 2. Estimated locations of elephant tissue and fecal samples from across Africa when assignments are allowed to vary anywhere within the elephants' range. All tissue and scat samples (*n* = 399) successfully amplified at seven or more loci. Sampling locations are indicated by a cross and are color-coded according to actual broad geographic region of origin: West Africa, Central forest, and Central, South, and East savannah (color-coded as in the Fig. 1 legend). Assigned location of each individual sample is shown by a circle and is color-coded according to its actual region of origin. The closer each circle is to crosses of the same color, the more accurate is that individual's assignment.

supporting information on the PNAS web site). However, it shares perhaps the biggest drawback of the SAM for our application: the implicit assumption that each sample comes from one of our 28 sampling locations. This assumption may not hold for tusks of unknown origin. Therefore, perhaps the greatest advantage of our smoothing method for estimating allele frequencies is that it leads to our CAM that does not make this assumption, but allows that the sample may have arisen from any location in the elephant's range. Fig. 2 shows the estimated location of each sample, color-coded according to actual broad geographic region of origin: West Africa, Central forest, and Central, South, and East savannah, using the CAM. Table 3 summarizes the distances between estimated location and actual sampling location of each individual, within each

region. Overall, 50% of samples were located to within 499 km of their actual origin, and 80% to within 932 km. Accuracy was greatest for samples from Central and Western regions (Table 3, see top row per region). This result might have been expected for the forest populations because they exhibit considerably greater genetic diversity than their savannah counterparts (14), and their habitat seems likely to produce greater barriers to gene flow (27). The high accuracy for central savannah elephants is presumably due to isolation from their East and South counterparts by long distances of forest and desert habitat. In contrast, the larger discrepancies for some individuals from East and Southern Africa reflect greater genetic similarity between these regions, possibly due to relatively few barriers to gene flow in these savannah habitats (28).

The CAM also estimates how much confidence to place in assigned sample origins by giving a set of plausible locations for any sample. For example, Fig. 3*A* shows 100 plausible locations for samples from Bia, ET, and Mikumi. For each sample, these 100 plausible locations are drawn from the set of all possible locations, weighted according to their probability. The tight clustering of the points corresponding to the Bia sample indicate high confidence that the sample comes from near Bia, the intermediate clustering for the ET sample indicate slightly less certainty, whereas the higher dispersion for the Mikumi sample indicates that we could not be confident of its precise origin.

Finally, we examined how the CAM might perform for samples from locations not included in our reference database. To perform this assessment, we applied it to each sample in turn, but excluded all samples that were from the same location as the sample being assigned when estimating allele frequencies. As expected, this results in a decrease in average assignment accuracy (Table 3, bottom row per region). For some locations, such as ET (and Garamba, data not shown) the effects of removing the samples from that location are particularly strong because the data available at other locations do not allow accurate allele frequency estimates to be obtained for these populations. For other locations, such as Mikumi, the effects of removing the samples at that location are small because samples at nearby locations provide adequate allele frequency estimates. This finding is illustrated in Fig. 3*B*, which shows the effect of excluding all other samples from a location on the confidence in assignments for the samples previously shown in Fig. 3*A*. The confidence in the assignment of the sample from Mikumi is barely affected. The results for the sample from Bia in Fig. 3 illustrate the complex structure of forest populations: when Bia samples are included in the assignment, the sample is precisely assigned to Bia, indicating that Bia is somewhat genetically distinct
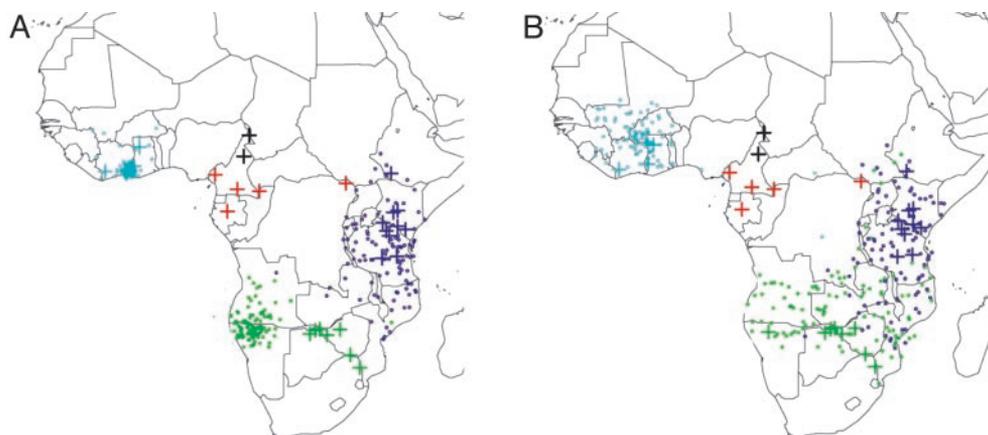
Fig. 3. Representation of confidence of assignments when all neighboring samples from that subpopulation (represented by a single cross) were included (*A*) vs. excluded (*B*) from the calculation of geographic-specific allele frequencies for ET (green), Bia (cyan), and Mikumi (blue). The 100 color-coded circles are random draws from the set of all possible locations weighted according to their probability. The concentration of these 100 circles in any given area gives a guide to the probability that the sample arose from that area under each condition.

from nearby Mole; when Bia is removed the precision of the assignment decreases, although the sample is still confidently assigned to West Africa. Collectively, this result indicates that the sample is more similar to the other samples from Mole than to samples from Central forest regions.

Discussion

Some of the most pressing needs in elephant conservation include timely identification of current poaching "hot spots" where greater law enforcement is needed; monitoring impacts of international trade decisions on elephant poaching throughout the African continent; determining whether declared government stockpiles are being illegally traded and replenished; determining whether sanctioned one-time sales include nonsanctioned tusks originating from other locations; and determining whether stockpiles of illegal ivory across Africa are being consolidated and exported. The accuracy with which genetic methods can determine the origin of DNA isolated from small amounts of ivory, tissue, or scat could greatly contribute to each of the above elephant conservation and management needs.

Simply discriminating between trade in forest and savannah elephant ivory should have considerable management significance in light of genetic evidence supporting the reclassification of forest and savannah elephants as separate species (14, 23–25). However, the genetic differentiation apparent among forest blocks (Table 1) suggests that we may be able to discern individual forests in which poaching is most heavily concentrated. These data may also assist authorities in tracking the bush meat trade, which increasingly includes elephant meat (8). Subanalyses of smaller tusks may be particularly useful here; unlike the ivory trade, bush meat poaching tends to occur independent of tusk size. Determining the proportion of African vs. Asian elephant ivory being sold in markets throughout the world is another important, and feasible use of these techniques (14).

A tremendous amount of information could be obtained simply by genetically monitoring the geographic composition of seizures as well as that of ivory moving through the major world ivory markets over time (e.g., Africa: Abidjan, Lagos, Dakar, Cairo, and Addis Ababa; Asia: Bangkok, Guangzhou, Tokyo, and Osaka; E. Martin, personal communication; see also ref. 6). Such information could greatly assist CITES by complementing efforts of the Monitoring Illegal Killing of Elephants (MIKE) (29) and Elephant Trade Inventory System (ETIS) (30) groups and the at a substantially reduced cost. The estimated cost of MIKE alone is $13.8 million over 6 years (29). These costs will only increase as populations continue to decline and elephants become increasingly difficult to detect. By contrast, DNA analyses are relatively inexpensive (currently approximately $100 per tusk, plus labor), likely to decrease with time, and reference sites do not require resampling over time.

Our methods make a number of modeling assumptions, some of which may hold only approximately in practice. However, the results in Figs. 2 and 3 and Tables 1–3 are based on the empirical performance of the method on samples of known origin, and are therefore valid independent of modeling assumptions, and indeed of other factors such as genotyping error and nonamplification of alleles.

The fact that accuracy is considerably greater for samples from locations in our reference database (Table 3) demonstrates the crucial importance of obtaining samples from regions not currently represented in our data. Whereas the considerable structure apparent in our results, particularly in forest regions (Table 1), holds great promise for genetic methods to determine sample location rather precisely, denser sampling will be required for this information to be fully exploited. Use of noninvasive means of acquiring DNA from feces should greatly facilitate these efforts, and the spatial smoothing used by our CAM means that relatively small numbers of samples from each location should suffice. This idea is particularly important when using samples such as scat where amplification success is uneven.

In conclusion, considerable breakthroughs in wildlife conservation and management should result from the methods we describe to track the geographic origin of poached ivory. Nearly all of the applications suggested above can be addressed with our existing data, although a few may require more reference samples from strategic areas to realize their full potential. The ability to acquire DNA from scat, along with tools to facilitate its collection over large remote areas (9), can also make such methods achievable in a timely manner for other species throughout the world.

The methods described here are implemented in the software package SCAT (Smoothed and Continuous Assignment Tests), which can be accessed at www.stat.washington.edu/stephens/software.html.

1. Dublin, H., Milliken, T. & Barnes, R. F. W. (1995) *Four Years After the CITES Ban: Illegal Killings of Elephants, Ivory Trade and Stockpiles* (International Union for Conservation of Nature and Natural Resources, Gland, Switzerland).
2. Robinson, J. G. &Bennett, E. L. (1999) *Hunting for Sustainability in Tropical Forests* (Columbia Univ. Press, New York).
3. Baker, C. S., Cipriano, F. & Paulmbi, S. R. (1996) *Mol. Ecol.* **5,** 671–685.
4. Douglas-Hamilton, I. (1987) *Pachyderm* **8,** 1–10.
5. Said, M. Y., Change, R. N., Craig, G. C., Thouless, C. R., Barnes, R. F. W. & Dublin, H. T. (1995) *African Elephant Database* (International Union for Conservation of Nature and Natural Resources, Gland, Switzerland).
6. Martin, E. & Stiles, D. (2000) *The Ivory Markets of Africa* (Save the Elephants, Nairobi).
7. Douglas-Hamilton, I. (1988) *African Elephant Database Project: Phase Two, December* (United Nations Environment Programme, Nairobi).
8. Courable, M., Hurst, F. & Milliken, T. (2003) *More Ivory Than Elephants: Domestic Ivory Markets in Three West African Countries* (Traffic International, Cambridge, U.K.).
9. Wasser, S. K., Houston, C. S., Koehler, G. M., Cadd, G. G. & Fain, S. R. (1997) *Mol. Ecol.* **6,** 1091–1097.
10. Wasser, S. K., Davenport, B., Ramage, E. R., Hunt, K. E., Parker, M., Clarke, C. & Stenhouse, G. (2004) *Can. J. Zool.* **82,** 475–492.
11. Comstock, K. E., Ostrander, E. A. & Wasser, S. K. (2003) *Cons. Biol.* **17,** 1–4.
12. Paetkau, D., Calvert, W., Sterling, I. & Strobeck, C. (1995) *Mol. Ecol.* **4,** 347–354.
13. Comstock, K. E., Wasser, S. K. & Ostrander, E. A. (2000) *Mol. Ecol.* **9,** 1004–1006.
14. Comstock, K. E., Georgiadis, N., Pecon-Slattery, J., Roca, A. L., Ostrander, E. A., O'Brien, S. J. & Wasser, S. K. (2002) *Mol. Ecol.* **11,** 2489–2498.
15. Nyakaana, S. & Arctander, P. (1998) *Mol. Ecol.* **7,** 1436–1437.
16. Karesh, W., Smith, F. & Frazier-Taylor, H. (1989) *Cons. Biol.* **1,** 261–262.
17. Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L. P. & Bouvet, J. (1996) *Nucleic Acids Res.* **24,** 3189–3194.
18. Fernando, P., Vidya, T. N., Rajapakse, C., Dangolla, A. & Melnick, D. J. (2003) *J. Hered.* **94,** 115–123.
19. Anderson, E. C. & Thompson, E. A. (2000) *Genetics* **16,** 1217–1229.
20. Diggle, P. J., Moyeed, R. A. & Tawn, J. A. (1998) *Appl. Stat.* **47,** 299–350.
21. Vounatsou, P., Smith, T. & Gelfand, A. E. (2000) *Biostatistics* **1,** 177–189.
22. Cornuet, J. M., Piry, S., Luikart, G., Estoup, A. & Solignac, M. (1999) *Genetics* **153,** 1989–2000.
23. Roca, A. L., Georgiadis, N., Pecon-Slattery, J. & O'Brien, S. J. (2001) *Science* **293,** 1473–1475.
24. Barriel, V., Thuet, E. & Tassy, P. (1999) *Evolution (Lawrence, Kans.)* **322,** 447–454.
25. Grubb, P., Groves, C. P., Dudley, J. P. & Shoshani, P. (2000) *Elephant* **2,** 1–4.
26. Whyte, I., Biggs, H., Gaylord, A. & Brack, L. (1997) in *Management Plan for the Kruger National Park: Policy Proposals Regarding Issues Relating to Biodiversity Maintenance, Maintenance of Wilderness Qualities, and Provision of Human Benefits*, (South African National Parks, Pretoria, South Africa), Vol. III, pp. 117–139.
27. Eggert, A, Rasner, C. A. & Woodruff, D. S. (2002) *Proc. R. Soc. London Ser. B* **269,** 1993–2006.
28. Georgiadis, N., Bischof, L., Templeton, A., Patton, J., Karesh, W. & Western, D. (1994) *J. Hered.* **85,** 100–104.
29. *CITES Document CoP12 Doc.34.2* (CITES, Santiago, Chile).
30. Traffic Report (2000) *Elephant Trade Inventory System (ETIS)* (Traffic International, Cambridge, U.K.).