DOI:10.1021/jf803680c

JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY
ARTICLE

Universal DNA-Based Methods for Assessing the Diet of Grazing Livestock and Wildlife from Feces

Anthony Pegard,[†] Christian Miquel,[†] Alice Valentini,[†] Eric Coissac,[†] Frédéric Bouvier,[§] Dominique François,[#] Pierre Taberlet,[†] Erwan Engel,[⊥] and François Pompanon*,[†]

[†]Laboratoire d'Ecologie Alpine, CNRS UMR 5553, Université Joseph Fourier, B.P. 53, 38041 Grenoble Cedex 9, France, [§]INRA, UE332, Domaine de la Sapinière, 18390 Osmoy, France, [#]INRA UR631 Amélioration Génétique des Animaux, Chemin de Borde-Rouge—Auzeville, B.P. 52627, 31326 Castanet-Tolosan Cedex, France, and [⊥]INRA, UR370 Qualité des Produits Animaux, 63122 Saint-Genès-Champanelle, France

Because of the demand for controlling livestock diets, two methods that characterize the DNA of plants present in feces were developed. After DNA extraction from fecal samples, a short fragment of the chloroplastic *trn*L intron was amplified by PCR using a universal primer pair for plants. The first method generates a signature that is the electrophoretic migration pattern of the PCR product. The second method consists of sequencing several hundred DNA fragments from the PCR product through pyrosequencing. These methods were validated with a blind analysis of feces from concentrate- and pasture-fed lambs. The signature method allowed differentiation of the two diets and confirmed the presence of concentrate in one of them. The pyrosequencing method allowed the identification of up to 25 taxa in a diet. These methods are complementary to the chemical methods already used. They could be applied to the control of diets and the study of food preferences.

KEYWORDS: Sheep; DNA barcoding; diet analysis; grazing livestock; feces; *trn*L intron; pyrosequencing; capillary electrophoresis

INTRODUCTION

Several factors such as recent food scares (e.g., bovine spongiform encephalopathy or avian flu), the use of genetically modified organisms (GMO) for feeding livestock, or religious reasons have reinforced public awareness regarding the origin of food products (*1*, *2*). This results in increased consumer demand for strict specifications on animal breeding that can be guaranteed by quality labels. Certifying husbandry conditions assumes that analytical tools exist and can authenticate the processes used, especially the animal diet.

Several methods have been developed that aim to specify the diet of livestock species. They are based on the fact that the diet influences the composition of the animal's end products, such as meat or milk. Therefore, specific compounds for which the nature or proportion is specific to a diet are used as diet markers. These compounds are either directly transferred from the plant eaten or produced by the animal's metabolism or ruminal microorganisms. They are mainly carotenoids (*3*), polyphenols (*4*), fatty acids (*5*), volatile compounds (*6*), and the ratios of stable isotopes (*7*). Moreover, a signature characterizing the global composition of meat or milk can also be used. For example, the global analysis of volatile compounds (*8*) and the spectral characterization of samples through near-infrared reflectance spectroscopy (NIRS; (*9*)) have already been used for diet authentication. Using these methods, the diet can be controlled from the early steps of husbandry (e.g., by analyzing feces) until the sale (e.g., by analyzing end products). However, these methods allow discrimination between different diets such as grass versus hay or ensilage (e.g., see refs (*10*−*12*)) but cannot give precise information about diet composition. Thus, complementary methods are necessary to provide a reliable and accurate identification of plants species that are eaten. These methods involve the analysis of gut contents or feces. The analysis of feces is attractive because it is a noninvasive method that can be realized through different techniques. The microscopic examination of plant cuticle fragments (e.g., ref (*13*)) is one of the most commonly used methods, but it requires extensive training, and a variable number of plant fragments remain unidentifiable. The chemical analysis of natural alkanes of plant cuticular wax (*14*) or the spectrum obtained by NIRS is also used for feces analysis, but sometimes the identification of plant species from complex mixtures is difficult.

A more accurate method for characterizing diet composition is the identification of the DNA fragments from the plant residues remaining in the feces. Recent techniques allow species identification based on the amplification and analysis of DNA even from degraded organic substrates (see ref (*15*) for a review). They are applied in forensics (*16*, *17*), to the analysis of fossils (*18*), to ecology (*19*), and in the food industry (*20*). Species identification can be set up using sets of specific primer pairs, each pair amplifying a single species or group of closely related species. The presence/absence of each species or group is then detected by

*Corresponding author (telephone +33 476 63 56 99; fax +33 476 51 42 79; e-mail francois.pompanon@ujf-grenoble.fr).

 Published on Web 06/16/2009

the success of the corresponding PCR amplification (e.g., ref (*21*)). Universal primer pairs have also been used, allowing the amplification of a given DNA fragment for a large set of species in a single PCR (*21*, *22*). Species identification is then possible by analyzing the variability of the fragment amplified. Taberlet et al. (*23*) recently designed a pair of primers targeting the P6 loop of the chloroplast *trn*L (UAA) intron. This fragment is adequate for the identification of DNA remaining in feces because the primers are universal in plants (i.e., highly conserved for angiosperms and gymnosperms), and the short size of the target fragment (10−143 base pairs without priming sites) allows the study of degraded DNA. Furthermore, the identification of plants is efficient because the amplified region is one of the most variable systems in size and sequence known to date. This system has already been shown to be relevant for analyzing the diet of wild herbivorous species including mammals, birds, molluscs, and insects (*24*).

In this paper, we describe two universal methods based on the analysis of the chloroplast *trn*L intron (*trn*L approach) from feces for characterizing the diet of livestock species. By analyzing the DNA present in lamb feces that were fed two different diets, we show how the *trn*L approach can be used either to quickly produce a specific signature of the diet or to identify the plant species eaten.

MATERIALS AND METHODS

Blind Analysis. All collected feces were given an anonymous reference to be blindly analyzed. After DNA extraction and PCR amplification of the trnL region, the variability of the PCR product was revealed by two methods. Its pattern of migration was produced by capillary electrophoresis [i.e., fragment length analysis (*25*)], and the amplicons were sequenced by massive parallel pyrosequencing. The samples were blindly managed during all of these steps, and the anonymity was raised after the production of the migration pattern or after the identification of plant taxa corresponding to the amplicons sequenced.

Lamb Diet and Sampling. A herd of 14 male lambs born during a 2 week period were divided into two groups of 7 after weaning at the target age of 45 days. Each group was then fed the same diet for the following 165 days. The first group ("pasture" diet, P) was exclusively pasture fed (pasture mainly composed of dactyl) with no additional feed, whereas the second group ("sheepfold" diet, S) was kept indoors and fed a pelleted mixed concentrate with 30% hay. The pelleted mixed concentrate (Agno Finition INRA) was composed of 33% wheat (*Triticum aestivum*), 32% beet (*Beta vulgaris*) pulp and molasses, 15% barley (*Hordeum vulgare*), 5% rapeseed (*Brassica napus*), 5% corn (*Zea mays*), and 3% soy (*Glycine max*). For each lamb, fresh fecal samples were collected and conserved in 96% ethanol until DNA extraction. Moreover, five samples of pelleted mixed concentrate ("concentrate", C) were collected.

Samples of all of the plant species composing the concentrate have been collected to obtain their migration pattern under capillary electrophoresis (*trn*L signature).

DNA Extraction. All extractions were performed in a room dedicated to nucleic acids extraction. Total DNA was extracted from about 20 mg of ground sheep feces with the DNeasy Tissue Kit (Qiagen GmbH, Hilden, Germany) following the manufacturer's protocol. For plants and concentrate, the same quantity of sample was processed with the DNeasy Plant Mini Kit (Qiagen) following the manufacturer's protocol.

***trn*L Signature of the Diet.** The amplification of the P6 loop of the *trn*L intron with the primers g (5′-GGGCAATCCTGAGCCAA-3′) and h (5′-CCATTGAGTCTCTGCACCTATC-3′) (*23*) generated a "+A" artifact (*26*, *27*), which would affect the migration pattern in capillary electrophoresis by producing two types of fragments from a single template (i.e., amplicons identical to the template or with a supplementary A). To avoid this artifact, a TCC pigtail (*26*) has been added to the 5′ end of the g primer (i.e., g* primer, 5′-TCCGGGCAATCCTGAGCCAA-3′), the PCR was performed with no extension step (*27*), and the PCR products were immediately stored at −20 °C. The g* and h primers were labeled with hexachloro-6-carboxyfluorescein (HEX) and carboxyfluorescein (6FAM), respectively, to allow simultaneous fluorescent detection during capillary electrophoresis. The amplification reactions were made in a final volume of 25 μL containing 2.5 μL of 10×x PCR buffer II (Applied Biosystems, Foster City, CA), 2 mM $MgCl_2$, 0.1 mM of each dNTP, 0.3 μM of each primer, 20 mg/mL of bovine serum albumin (BSA, Roche, Basel, Switzerland), 0.6 U of AmpliTaq Gold Polymerase (Applied Biosystems), and 2.5 μL of 500× diluted DNA extract (i.e., 0.1 ng of template DNA). The dilution of the DNA extract avoided high rates of PCR inhibitors. The mix was realized with UHQ water exposed to UV rays (5 J/cm^2) to eliminate DNA contaminants. The amplification reaction consisted of 10 min at 95 °C followed by 45 cycles of 30 s at 95 °C and 30 s at 55 °C.

The polymorphism of the migration pattern of the PCR products, which depends on the size and sequence of each DNA fragment, was checked for each strand of the amplicon under denaturing conditions. For each sample, 1 μL of 5× diluted PCR products was added to 10 μL of formamide and 0.2 μL of ROX350 (size standard) and electrophoresed for 35 min on an ABI PRISM3130 Genetic Analyzer (Applied Biosystems) using 36 cm capillaries and POP-7 polymer. Fluorograms were analyzed using Genemapper3.7 software package (Applied Biosystems).

Identification of the Species. To evaluate the efficiency of new sequencing methods with regard to the classical methods (i.e., cloning and sequencing), six previously studied samples (two fecal samples for each diet and two concentrates) were analyzed following both strategies.

Cloning and Sequencing. One sample representative of each type of migration pattern was chosen to perform a new PCR with nonfluorescent g and h primers. The PCR product has been cloned into PCR 2.1-TOPO vector, TOPO TA Cloning kit (Invitrogen, Burlington, ON, Canada) following the manufacturer's protocol, except that the ligation was stored at 4 °C for 2 days before heat shocking competent *Escherichia coli* for 45 s at 42 °C. After 24 h of culture on LB medium plates at 37 °C, 192 clones per PCR product were picked up with a pipet tip and put in 30 μL of sterile water. We amplified 2.5 μL of this solution by PCR using the M13(-20)-forward and M13-reverse primers. PCR products were purified with the Qiaquick PCR Purification Kit (Qiagen) according to the manufacturer's protocol. Before the clones were sequenced, a prescreening was done by amplifying the *trn*L region with the g*/h primer pair and revealing the migration pattern of the purified PCR product under capillary electrophoresis. Then, we sequenced two of the clones exhibiting the same migration pattern for a given category of sample. Both DNA strands were sequenced using the Big Dye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems) in 20 μL containing 30 ng of purified DNA and 4 pmol of M13 amplification primers according to the manufacturer's specifications. Sequencing reactions underwent 25 cycles of 30 s at 96 °C, 30 s at 50 °C, and 4 min at 60 °C. Excess dye terminators were removed by Sephacryl/Sephadex column purification. Sequencing reactions were electrophoresed for 45 min on an ABI PRISM 3100 Genetic Analyzer (Applied Biosystems) using 36 cm capillaries and POP-7 polymer. Sequences were analyzed using BioEdit [version 7.0.5.2 (*28*)].

Massive Pyrosequencing. A new PCR was carried out for each of the six samples analyzed. Because PCR products from all samples had to be mixed before massive sequencing, we had to tag them to find the origin of the sequences obtained after pyrosequencing. Thus, for each sample the PCR was carried out with the g and h primers to which a different (sample-specific) six-nucleotide sequence has been added at the 5′ end (*24*). These tags allowed assigning each sequence obtained to its sample. PCR conditions were identical to those previously described with the primers g* and h. The PCR products were purified using the MinElute PCR purification kit (Qiagen), and DNA was quantified with a NanoDrop ND-1000 UV−vis spectrophotometer (NanoDrop Technologies, Wilmington, DE). Then the PCR products from the different samples were mixed in equimolar proportions and sequenced by large-scale pyrosequencing on a 454 Genome Sequencer FLX (Roche, Basel, Switzerland) at the Genoscope (Centre National de Séquençage, Evry, France), following the manufacturer's instructions.

Setting up Reference Databases. The sequences produced were identified by comparison to those present in a reference database. This database (Gbpln database) was established as a subset of the plant division of GenBank, by extracting the g−h region of the *trn*L P6 loop from GenBank sequences using the ecoPCR software (freely available at http://www.grenoble.prabi.fr/trac/ecoPCR). This consisted of a virtual PCR with the g and h primers, allowing up to three mismatches between each

primer and the sequence amplified and keeping only the sequences having a length of between 20 and 500 base pairs (fragment length without primers).

Data Analysis. All of the sequences produced by the massive parallel pyrosequencing were sorted out from the outfile using the sample-specific tag. For each sample, we selected for further analyses the sequences that were repeated at least three times to avoid the influence of sequence errors (*30*). We also removed sequences with an error within the tag sequence because this prevented the assignment of the sequence to a sample without ambiguity. Whatever the sequencing method, we compared each sample sequence to those of the reference database using the ecoTag software (freely available at https://www.grenoble.prabi.fr/svn/OBISofts/OBITools/tags/OBITools-1.0.0/src/). This program, which was developed in our laboratory, worked in two steps. The first one used the FASTA algorithm (*31*) to align each sequence from the sample (i.e., query sequence) to the sequences of the database and calculated their similarity. For each query sequence the program selected the sequences of the database with at least 98% of identity on the whole length (100%) of the query sequence. Each sequence of the Gbpln database was assigned to a taxon labeled by its taxon identifier (taxid) following the NCBI taxonomy (*32*, *33*). Then, the program assigned to each query sequences all of the taxids of similar sequences found in the database. The second step consisted of assigning a unique taxon to the query sequence. This unique taxon corresponded to the last common ancestor (in the NCBI taxonomic tree) of all the taxids previously retained. For example, if a query sequence corresponded to one or several sequences with the same taxid in the database (e.g., *Pisum sativum*), it was identified as a sequence of the corresponding species (*P. sativum*). Conversely, if a query sequence corresponded to several sequences with different taxids corresponding to several species (e.g., *Brassica napus*, *Eruca sativa*, etc.), the taxon assigned to the query sequence was the lower rank taxon common to all of these species (family Brassicaceae in this example).

RESULTS

***trn*L Signature of the Diet.** For each type of sample (i.e., pasture diet, sheepfold diet, and concentrate), all replicates (i.e., 7, 7, and 5, respectively) exhibited the same migration patterns (Figure S1 of the Supporting Information). **Figure 1** gives one migration pattern for each kind of sample. The patterns obtained for different types of samples were clearly different, showing that a specific signature could characterize each category. Moreover, combinations of peaks specific of the concentrate's signature were found in the signatures of the sheepfold samples, with different relative intensities. This showed the presence of concentrate in the sheepfold diet (**Figure 1**). The analysis of the migration pattern of the PCR products obtained from pure plant extracts (data not shown) showed that these peaks corresponded to rapeseed, barley, and wheat or soy. However, the fact that several species such as maize and soy showed closely matched PCR products (same fragment size and close DNA sequence) and could not be identified by their migration pattern prevented an unambiguous identification of the species eaten. Thus, although this technique gave a specific signature of each diet, a peak in the signature did not always correspond to a single species.

Identification of the Species Eaten. *Cloning and Sequencing*. For each of the 6 PCR products (2 replicates per category of sample) the prescreening of the 96 clones revealed 8, 5, and 3 different sizes of insert for the pasture diet, sheepfold diet, and concentrate extracts, respectively. Two clones with the same size of insert have been sequenced per sample. In all cases inserts of the same size had the same DNA sequence. This strategy allowed the identification of up to 8 taxa per diet (**Table 1**).

Massive Pyrosequencing. The sequencing of the 2 replicates per category of sample gave 27380 sequences that corresponded to a mean ± SD of 2282 ± 1152 of amplicons sequenced per sample. Once the sequences with errors have been removed (see Materials and Methods), a mean number of 1867 ± 880 sequences per sample remained. For a given sample, the most represented sequence occurred between 1012 and 227 times. The DNA of beet and corn was not or hardly detectable in the concentrate. The other concentrate compounds (i.e., wheat, barley, rapeseed, and soy) were detectable in concentrate extracts as well as in feces of lambs with the sheepfold diet (**Table 1**). Many sequences of Triticeae were found in all sample categories because wheat (*Triticum aestivum*) shared the same *trn*L intron sequence with several other Poaceae. Twenty-five, 22, and 13 taxa were identified as composing the pasture diet, the sheepfold diet, and the concentrate extracts, respectively (**Table 2**). They corresponded to 46 different taxa that were identified at the species level (24.4%), the genus level (51.1%), or the family level (91.1%). Most of the taxa composing the diets were from the families Poaceae, Fabaceae, and Polygonaceae.

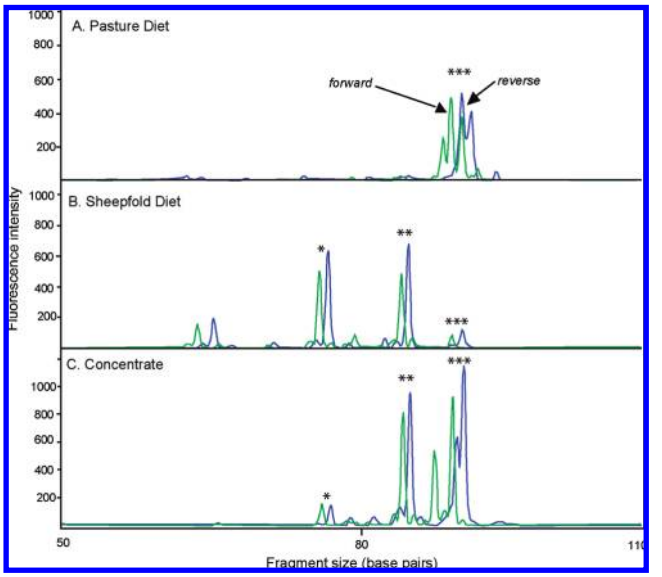

Figure 1. *trn*L signature of two feces of lambs with different diets and of a food sample. Each signature is the multipeak profile obtained after capillary electrophoresis in denaturing conditions of the fluorescent PCR products obtained using the *g* and *h* primers. Each signature is representative of all samples from the same category: feces from sheep with the pasture diet (**A**), feces from sheep with the sheepfold diet (**B**), and concentrate INRA Agno Finition (**C**). Migration patterns are figured for both DNA strands (forward and reverse). Each peak represents at least one plant species. Peaks: ∗, rapeseed; ∗∗, barley; ∗∗∗, wheat and soy.

DISCUSSION

The power of the two DNA-based methods developed for assessing the diet of grazing livestock mainly resides in the characteristics of the fragment analyzed. With the massive sequencing approach, its short size and the high variability between conserved priming sites allow the characterization of the DNA of all plant species that are present in degraded substrates. The two methods have complementary applications. The *trn*L signature that is based on capillary electrophoreses is quick, accessible, and robust. The migration pattern obtained allows detecting a combination of peaks characteristic of a complex component (here the concentrate) within a diversified diet (**Figure 1**). However, this method is not efficient for identifying the whole diet composition (**Table 2**) because several species can account for the same peak. The second method uses massive pyrosequencing and requires a new generation genome sequencer. Nevertheless, it is an accessible technique because the protocol is simple enough to be implemented in a laboratory used to

Table 1. Identification of Plant Taxa Present in a Food Concentrate Extract and in Feces of Lambs Submitted to Two Different Diets[a]

			no. of replicates (over 2) in which the taxon was present (pyrosequencing−cloning approaches)		
family	taxon identified	taxon level	pasture diet	sheepfold diet	concentrate
	Magnoliophyta	no rank	1−0		
	Malpighiales	order		2−0	
Amaranthaceae	*Atripex patula*	species		1−0	
Asteraceae	Asteraceae group 1	family		1−1	2−0
	Asteraceae group 2	family	1−0	1−0	
	Cymbonotus lawsonianus	species		2−0	
Chenopodiaceaeae	***Beta vulgaris* (beet)**	species		1−0	
Clusiaceae	*Hypericum kamtschaticum*	species	2−0		
Convolvulaceae	Convolvulaceae	family	1−1	1−0	
Brassicaceae	Brassicaceae	family		2−1	2−0
Fabaceae	Papillionoideae group1	subfamily		2−0	
	Papillionoideae group 2	subfamily		2−0	
	Medicago sativa	species		1−1	1−0
	Lotus sp.	genus	2−1		
	Lathyrus pratensis	species	1−0		
	Trifolium sp.	genus	1−0		
	Vicia faba	species		2−0	
	Lupinus sp.	genus		1−0	
	Pisum sativum	species			2−0
	***Glycine max* (soy)**	species		2−0	2−2
Geraniaceae	Geraniaceae	family	1−1		
Juglandaceae	Juglandaceae	family	2−1		
Lamiaceae	*Ajuga multiflora*	species	1−0		
Poaceae	Pooideae group1	subfamily			2−0
	Pooideae group 2	subfamily			1−0
	Pooideae group 3	subfamily	2−0		1−0
	Pooideae group 4	subfamily	2−0		
	Pooideae group 5	subfamily	1−0		
	Agrostis sp.	genus	2−0		
	Poaceae group 1	family	0−1		1−0
	Poaceae group 2 (including corn)	family			1−0
	Bromus sp.	genus	1−0		
	Poeae	tribe	1−0		
	Avena sativa	species		1−0	1−0
	Festuca scariosa	species			1−0
	***Triticeae* (including wheat)**	tribe	2−1	2−1	2−1
	***Hordeum* sp. (barley)**	genus	2−1	2−1	2−2
Polygonaceae	Polygonaceae group 1	family		1−0	
	Polygonaceae group 2	family		1−0	
	Polygonaceae group 3	family	1−0		
	Polygonum sp.	genus	2−0	1−0	
	Rumex sp.	genus	2−0	1−0	
	Fallopia sp.	genus		2−0	
Rosaceae	*Potentilla* sp.	genus	2−1		
	Rosoideae	subfamily	1−0		
	Rubus sp.	genus	1−0		
Verbenaceae	*Verbena* sp.	genus	2−0		

[a] For each taxon, the table gives the number of sample replicates (over two) in which the taxon was detected according to the pyrosequencing and cloning approaches, respectively. Plant taxa were identified by comparison to the "GbpIn database" built from the GenBank plant section. Different groups mentioned for the same taxon (for example, Pooidae groups) refer to different sets of sequences identified at the same taxonomic level. Taxa in bold correspond to the component of the concentrate.

Table 2. Comparison of the DNA-Based Methods Used for Diet Identification

		*trn*L signature	cloning/sequencing	massive pyrosequencing
ability to identify concentrate in the diet		yes	yes	yes
no. of taxa identified	pasture diet	1	8	25
	sheepfold diet	3	5	22
	concentrate	3	3	13
no. of exptl steps/sample		1 PCR, 1 CE[a]	1 PCR, 1 cloning, 96 CE, ≤16 sequences	1 PCR, 1/400 pyrosequencing run
universality of reference database		no, depending on the CE device	yes, established from public sequence databases	
approximate cost for consumable		10 euros/sample	650 euros/sample	30 euros/sample[b]

[a] CE, capillary electrophoresis. [b] Estimations given for a 454 Genome Sequencer FLX with titanium reagents.

work with degraded DNA, and several organizations offer the massive pyrosequencing of PCR products. Over the past year, the cost of this service has been divided by three, and it is now really competitive (**Table 2**). Moreover, this cost is expected to decrease again in the near future with the current development of sequencing technologies. The massive pyrosequencing is an efficient alternative of the cloning approach that was the only way to characterize the diversity of sequences composing a PCR product until now. It is cheaper and far less time-consuming for a much more comprehensive description of the diet (**Table 2**). With this approach, we identified up to 25 taxa in a diet, which was much more than the cloning approach and similar to the number of species identified in the sheep diet using samples collected from ruminal canulas (*34*). The straightforwardness of the protocol results from the use of universal primers that require a unique PCR amplification.

The classical techniques used for diet traceability such as terpenes or alkanes analyses and NIRS can discriminate among diets (*35*). For example, they allow the authentication of diets on the basis of the alternation of pasture and concentrate (*36*). Some techniques can also provide context-specific information, such as the terpene analysis that gives a signature of a pasture depending on the place and time of year (*5*). The main advantage of the massive sequencing approach with regard to the chemical analyses is its efficiency for identifying the major components of diets. The sequence identification relies on a universal (i.e., not place or time dependent) reference database containing the sequence of the studied fragment for all potentially eaten plants. For the *trn*L intron used for this study > 1500 species are already referenced in our laboratory and > 23000 sequences are available in GenBank. A limit of the *trn*L method is its low ability in discriminating species in some taxonomic groups. Typically, most of the Poaceae are only identified at the family level using *trn*L sequences. It is possible to overcome this drawback by combining loci that are complementary for species identification (*37*). Several loci already used in plant identification (e.g., refs (*38*) and (*39*)) could be used together with the *trn*L by multiplexing PCR products before massive pyrosequencing.

Contrary to DNA-based methods, chemical analyses can give insights into the characterization of the tissue eaten (e.g., root or leaf) or its state (e.g., grass or hay). For example, the amount of different fatty acids in lambs' muscles differs for diets based on grass or concentrate and hay (*10*). Thus, a complementary use of chemical and DNA-based methods can be anticipated (*24*). For instance, once the species eaten have been determined by analyzing DNA traces, the NIRS method has the potential to provide information about the tissue eaten (*9*) and the analysis of stable isotopes about the geographic origin (*40*). The DNA-based methods and chemical analyses are also complementary because the former identify the recent diet (a few days), whereas the latter give long-term information.

The sensitivity of the massive pyrosequencing approach can be evaluated by comparing the composition analysis of concentrate to that of feces of lambs partly fed with concentrate (sheepfold diet). The method is reliable because plant taxa that have been identified in all concentrate samples (5 over 6) have also been reliably identified in the feces. This was the case for soy, which composed only 3% of the concentrate, showing that the method has a low detection threshold. This has been confirmed by quantitative PCRs showing that a component can be identified in a complex substrate when its DNA represented at least 2% of the target DNA. Dilution did not affect this threshold because there were > 20 initial target-DNA copies (unpublished results). This low detection threshold would explain why we detected eight taxa that were not thought to be present in the concentrate composition according to the ingredients mixed. They may result from a contamination of ingredients used for making the concentrate. Because hay and concentrate were stored on the same site before being used for feeding lambs, plants from hay could have also contaminated the concentrate. This shows the high sensitivity of the method but also the risk of amplifying contaminants. This risk can be strongly reduced by limiting the number of PCR cycles (*24*). The detection threshold that may be generally low may also depend on the DNA contents of the food product. Rapeseed, which represents 5% of the concentrate compounds, was detected in both concentrate and feces, whereas corn and beet that represent 5 and 32% of the concentrate composition, respectively, were not detected in all of the concentrate extracts. This could be due to the very low concentration of DNA in some plant byproducts such as molasses. Due to this variation of DNA contents according to the tissue and/or its byproduct that is eaten, the proportion of sequences obtained by the pyrosequencing approach does not reflect the proportion of species in the diet. The digestibility of the plant eaten may also affect the quality and quantity of the target DNA present in the feces. However, the pasture diet is composed of leaves that may have similar digestibility and may contain similar amounts of chloroplastic DNA whatever the species. In this case the number of sequences obtained in a PCR product could reflect the relative frequency of its occurrence in the diet. This would be possible because the primer pair targets highly conserved regions across taxa (*23*) and thus limits the preferential amplification of a species (*24*). From a quantitative point of view, DNA-based analyses could be complementary to the alkane approach that can estimate the absolute quantity of plant eaten (*14*). However, the ability of DNA-based methods to reflect the relative quantities of species in the diet remains to be checked through suitable protocols with a strict control of ingested quantities.

We have shown that DNA-based analyses of feces can be used for a reliable and accurate assessment of the diet of grazing species. Such methods could be applied for the control and authentication of livestock diets. They can be used on samples collected at the farm, where the visual inspection of the food is not always informative. For instance, flocks with different certification levels (and thus different feeding requirements) may coexist in the same farm, and unexpected controls of feces composition could be informative. Together with the use of primers allowing plant detection described here, the use of animal-specific primers could allow the identification of the presence of animal byproducts in the diet. Such diet assessments could be performed concomitantly with the DNA characterization of the individual or breed that is possible from the same fecal sample (*41*). Among the DNA-based methods, the signature approach is convenient for the detection of a special diet component. This is also the case of PCR approaches using specific primers for certifying the presence/absence of target components such as GMO (e.g., ref (*42*)). The massive sequencing techniques using universal primers are best adapted to the global assessment of the diet and thus to the detection of even unexpected species. This last approach would be useful both for certification and research purposes on both livestock and wildlife. For example, it could advantageously replace behavioral studies or analyses of stomach contents for inferring food preferences.

ACKNOWLEDGMENT

We thank Patrik Mraz and Mihai Puscas for constructive advice and Carole Poillot, Delphine Rioux, and Ludovic Gielly for help with the experiments.

Supporting Information Available: Migration patterns of the PCR products. This material is available free of charge via the Internet at http://pubs.acs.org.

LITERATURE CITED

(1) Pascal, G.; Mahe, S. Identity, traceability, acceptability and substantial equivalence of food. *Cell. Mol. Biol.* **2001**, *47*, 1329–1342.

(2) Pinotti, L.; Moretti, V. M.; Baldi, A.; Bellagamba, F.; Campagnoli, A.; Savoini, G.; Cantoni, C.; Dell'Orto, V. Feed authentication as an essential component of food safety and control. *Outlook Agric.* **2005**, *34*, 243–248.

(3) Prache, S.; Theriez, M. Traceability of lamb production systems: carotenoids in plasma and adipose tissue. *Anim. Sci.* **1999**, *69*, 29–36.

(4) Besle, J.-M.; Lamaison, J. L.; Dujol, B.; Pradel, P.; Fraisse, D.; Viala, D.; Martin, B. Flavonoids and other phenolics in milk as a putative tool for traceability of dairy production systems. In *Indicators of Milk and Beef Quality*; EAAP Publication 112; Hocquette, J. F., Gigli, S., Eds.; Wageningen Academic Publishers: Wageningen, The Netherlands, 2005; pp 345−350.

(5) Engel, E.; Ferlay, A.; Cornu, A.; Chilliard, Y.; Agabriel, C.; Bielicki, G.; Martin, B. Relevance of isotopic and molecular biomarkers for the authentication of milk according to production zone and type of feeding of the cow. *J. Agric. Food Chem.* **2007**, *55*, 9099–9108.

(6) Sivadier, G.; Ratel, J.; Bouvier, F.; Engel, E. Authentication of meat products: determination of animal feeding by parallel GC-MS analysis of three adipose tissues. *J. Agric. Food Chem.* **2008**, *56*, 9803−9812

(7) Renou, J. P.; Bielicki, G.; Deponge, C.; Gachon, P.; Micol, D.; Ritz, P. Characterization of animal products according to geographic origin and feeding diet using nuclear magnetic resonance and isotope ratio mass spectrometry. Part II: Beef meat. *Food Chem.* **2004**, *86*, 251–256.

(8) Engel, E.; Ratel, J. Correction of the data generated by mass spectrometry analyses of biological tissues: application to food authentication. *J. Chromatogr., A* **2007**, *1154*, 331–341.

(9) Cozzolino, D.; De Mattos, D.; Martins, D. V. Visible/near infrared reflectance spectroscopy for predicting composition and tracing system of production of beef muscle. *Anim. Sci.* **2002**, *74*, 477–484.

(10) Aurousseau, B.; Bauchart, D.; Calichon, E.; Micol, D.; Priolo, A. Effect of grass or concentrate feeding systems and rate of growth on triglyceride and phospholipid and their fatty acids in the M-longissimus thoracis of lambs. *Meat Sci.* **2004**, *66*, 531–541.

(11) Martin, B.; Verdier-Metz, I.; Buchin, S.; Hurtaud, C.; Coulon, J. B. How do the nature of forages and pasture diversity influence the sensory quality of dairy livestock products? *Anim. Sci.* **2005**, *81*, 205–212.

(12) Priolo, A.; Prache, S.; Micol, D.; Agabriel, J. Reflectance spectrum of adipose tissue to trace grass feeding in sheep: influence of measurement site and shrinkage time after slaughter. *J. Anim. Sci.* **2002**, *80*, 886–891.

(13) Holechek, J. L.; Vavra, M.; Pieper, R. D. Botanical composition determination of range herbivore diets − a review. *J. Range Manage.* **1982**, *35*, 309–315.

(14) Dove, H.; Mayes, R. W. Plant wax components: a new approach to estimating intake and diet composition in herbivores. *J. Nutr.* **1996**, *126*, 13–26.

(15) Teletchea, F.; Maudet, C.; Hänni, C. Food and forensic molecular identification: update and challenges. *Trends Biotechnol.* **2005**, *23*, 359–366.

(16) Capelli, C.; Tschentscher, F.; Pascali, V. L. "Ancient" protocols for the crime scene? Similarities and differences between forensic genetics and ancient DNA analysis. *Forensic Sci. Int.* **2003**, *131*, 59–64.

(17) Wan, Q. H.; Fang, S. G. Application of species-specific Polymerase Chain Reaction in the forensic identification of tiger species. *Forensic Sci. Int.* **2003**, *131*, 75–78.

(18) Hofreiter, M.; Poinar, H. N.; Spaulding, W. G.; Bauer, K.; Martin, P. S.; Possnert, G.; Paabo, S. A molecular analysis of ground sloth diet through the last glaciation. *Mol. Ecol.* **2000**, *9*, 1975–1984.

(19) Ficetola, G. F.; Miaud, C.; Pompanon, F.; Taberlet, P. Species detection using environmental DNA from water samples. *Biol. Lett.* **2008**, *4*, 423–425.

(20) Maudet, C.; Taberlet, P. Holstein's milk detection in cheeses inferred from melanocortin receptor 1 (MC1R) gene polymorphism. *J. Dairy Sci.* **2002**, *85*, 707–715.

(21) Deagle, B. E.; Gales, N. J.; Evans, K.; Jarman, S. N.; Robinson, S.; Trebilco, R.; Hindell, M. A. Studying seabird diet through genetic analysis of faeces: a case study on macaroni penguins (*Eudyptes chrysolophus*). *PLoS ONE* **2007**, *2*, e831.

(22) Bradley, B. J.; Stiller, M.; Doran-Sheehy, D. M.; Harris, T.; Chapman, C. A.; Vigilant, L.; Poinar, H. Plant DNA sequences from feces: potential means for assessing diets of wild primates. *Am. J. Primatol.* **2007**, *69*, 699–705.

(23) Taberlet, P.; Coissac, E.; Pompanon, F.; Gielly, L.; Miquel, C.; Valentini, A.; Vermat, T.; Corthier, G.; Brochmann, C.; Willerslev, E. Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. *Nucleic Acids Res.* **2007**, *35*, 8.

(24) Valentini, A.; Miquel, C.; Nawaz, M. A.; Bellemain, E.; Coissac, E.; Pompanon, F.; Gielly, L.; Cruaud, C.; Nascetti, G.; Winker, P.; Swenson, J. E.; Taberlet, P. New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. *Mol. Ecol. Resour.* **2009**, *9*, 51–60

(25) Zinger, L.; Gury, J.; Alibeu, O.; Rioux, D.; Gielly, L; Sage, L.; Pompanon, F.; Geremia, R. A. CE-SSCP and CE-FLA, simple and high-throughput alternatives for fungal diversity studies. *J. Microbiol. Methods* **2008**, *72*, 42–53.

(26) Brownstein, M. J.; Carpten, J. D.; Smith, J. R. Modulation of non-templated nucleotide addition by tag DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* **1996**, *20*, 1004–1010.

(27) Magnuson, V. L.; Ally, D. S.; Nylund, S. J.; Karanjawala, Z. E.; Rayman, J. B.; Knapp, J. I.; Lowe, A. L.; Ghosh, S.; Collins, F. S. Substrate nucleotide-determined non-templated addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping and cloning. *Biotechniques* **1996**, *21*, 700–709.

(28) Hall, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **1999**, *41*, 95–98.

(29) Taberlet, P.; Gielly, L.; Pautou, G.; Bouvet, J. Universal primers for amplification of 3 noncoding regions of chloroplast DNA. *Plant Mol. Biol.* **1991**, *17*, 1105–1109.

(30) Huse, S. M.; Huber, J. A.; Morrison, H. G.; Sogin, M. L.; Mark Welch, D. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* **2007**, 8.

(31) Pearson, W. R.; Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 2444–2448.

(32) Benson, D. A.; Karsch-Mizrachi, I.; Lipman, D. J.; Ostell, J.; Rapp, B. A.; Wheeler, D. L. GenBank. *Nucleic Acids Res.* **2000**, *28*, 15–18.

(33) Wheeler, D. L.; Chappey, C.; Lash, A. E.; Leipe, D. D.; Madden, T. L.; Schuler, G. D.; Tatusova, T. A.; Rapp, B. A. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2000**, *28*, 10–14.

(34) Santos, G. R. D.; Batista, A. M. V.; Guim, A.; dos Santos, M. V. F.; Silva, M. J. D.; Pereira, V. L. A. Evaluation of botanical composition of sheep diet in Caatinga pasture. *Rev. Bras. Zootecn.* **2008**, *37*, 1876–1883.

(35) Vasta, V.; Priolo, A. Ruminant fat volatiles as affected by diet. A review. *Meat Sci.* **2006**, *73*, 218–228.

(36) Priolo, A.; Cornu, A.; Prache, S.; Krogmann, M.; Kondjoyan, N.; Micol, D.; Berdagué, J.-L. Fat volatiles tracers of grass feeding in sheep. *Meat Sci.* **2004**, *66*, 475–481.

(37) Kress, W. J.; Erickson, D. L. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. *PLoS ONE* **2007**, *2*, e508.

(38) Kress, W. J.; Wurdack, K. J.; Zimmer, E. A.; Weigt, L. A.; Janzen, D. H. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 8369–8374.

(39) Poinar, H. N.; Hofreiter, M.; Spaulding, W. G.; Martin, P. S.; Stankiewicz, B. A.; Bland, H.; Evershed, R. P.; Possnert, G.; Paabo, S. Molecular coproscopy: dung and diet of the extinct

ground sloth *Nothrotheriops shastensis*. *Science* **1998**, *281*, 402–406.
(40) Kelly, S.; Heaton, K.; Hoogewerff, J. Tracing the geographical origin of food: The application of multi-element and multi-isotope analysis. *Trends Food Sci. Technol.* **2005**, *16*, 555–567.
(41) Taberlet, P.; Camarra, J. J.; Griffin, S.; Uhres, E.; Hanotte, O.; Waits, L. P.; Dubois Paganon, C.; Burke, T.; Bouvet, J. Noninvasive genetic tracking of the endangered Pyrenean brown bear population. *Mol. Ecol.* **1997**, *6*, 869–876.
(42) Garay, A. G.; Mayas, I. G.; Garcia, G. V. Suitable method for simultaneous and specific detection of maize (*Zea mays* L.) and genetically modified soyabean (*Glycine max* L.) in animal feeds. *J. Anim. Feed Sci.* **2009**, *18*, 162–172.

Received November 26, 2008. Revised manuscript received April 28, 2009. Accepted May 22, 2009. This work has been possible through the collaborations set up within the "Authenticité Traçabilité" network (AUTRAC) from the French Institut National de la Recherche Agronomique (INRA). This study was partly supported by the French Agence Nationale pour la Recherche Project ANR-06-PNRA-024-06 and by the INRA AUTMAT research program. F.P. was supported by the INRA.