Accurate food animal identification is essential for improving disease control and enhancing food safety. The US Code of Federal Regulations (CFR, Title 9 Animals and Animal Products) presently requires individual animal identification and tracing of cattle with tuberculosis (§77.17) or brucellosis (§78.1), for traceback purposes. The Food Safety and Inspection Service (FSIS) also requires collection of animal identification for all cattle tested for bovine spongiform encephalopathy (BSE, FSIS notice 28-04). In beef packing plants, animals with antemortem or postmortem signs of disease are screened for antimicrobial, chemical, and drug residues that may violate food safety standards. Tissues from high-risk animals are tracked with a system of physical labels, the accuracy of which is critical for correctly identifying violations and removing condemned carcasses from the human food chain.

Current research suggests that DNA-based technology represents a promising means for verifying the accuracy of physical labels for identification of cattle. After ear tags and other physical identification devices have been removed, an animal’s DNA remains as a stable, accessible, integral, and identifiable component of its beef products. Thus, DNA-based identification technologies provide a gold standard for auditing the fidelity of physical labels. Historically, the feasibility of using DNA markers for identification of individual cattle has been limited by cost, availability, and demand. However, these constraints have been diminished by 3 recent events. First, the human genome project has resulted in a variety of efficient genotyping technologies that appear to be capable of producing single nucleotide polymorphism (SNP) genotypes for less than $0.01 each. Single nucleotide polymorphisms are common DNA sequence variations among individuals. Second, a set of well-char-
characterized, highly polymorphic SNPs in popular US cattle breeds is now available for use. Third, on December 23, 2003, the USDA announced the identification of a case of BSE in the United States, highlighting the immediate need for a national identification system for cattle.

When combined with physical, electronic, and database tools, DNA markers provide a means for verifying animal identification. The purpose of the study reported here was to determine whether a selected set of 20 SNP markers could be used to verify sample tracking in a commercial slaughter facility that processes primarily market dairy cows.

**Materials and Methods**

**Animals** — The study was conducted at a single large federally inspected commercial slaughter facility. Each week during a 4-week period in the summer of 2003, 42 animals were randomly chosen for inclusion in the study (168 total). Blood samples (9 mL) were collected from the tail vein of the selected cattle as they were driven to an enclosure for stunning. Matched liver samples (15 g) were collected during beef processing by conventional methods under which the position and sequence of each carcass was synchronized with its various beef-derived products as they moved down the processing line. Consistent with normal operational procedures of the plant, no tracking in a commercial slaughter facility that processes between beef and dairy cattle.

**DNA extraction and genotype analysis** — For blood samples, DNA was extracted by use of a solid-phase system incorporating 96-well microtiteration plates, according to the manufacturer's instructions. For liver samples, DNA was extracted as described. The DNA samples were randomly arrayed in 96-well microtiteration plates for genotype analysis. In addition, DNA from each of 48 randomly selected blood and liver samples was split and submitted in duplicate to estimate the discordance rate attributable to genotyping error.

A set of 20 di-allelic bovine SNP markers spanning 12 chromosomes was used. Markers were selected on the basis of availability for scoring with a commercial system by means of matrix-assisted, laser desorption-ionization, time-of-flight mass spectroscopy at a commercial contract laboratory that used assays and procedures similar to those described previously.

The DNA from a multibreed beef panel was used to compare SNP allele frequencies between beef and dairy cattle. This well-characterized multibreed panel (MARC beef cattle diversity panel, version 2.1) consisted of 92 sires from 16 popular beef breeds and 4 sires from the Holstein dairy breed. To maximize the total number of unshared haploid genomes, sires within each breed were selected for pedigrees with minimal relationships between ancestors. On the basis of the number of registered progeny for each breed, the beef breeds in this panel comprised >99% of the germplasm used in the US beef cattle industry.

**Mean probability of identity** (P$_I$) for an individual SNP marker was defined as the estimated probability that 2 unrelated individuals selected from the group at random would possess identical genotypes. Briefly, for an SNP marker (eg, locus A) with alleles A$_1$ and A$_2$, mean probability that 2 particular

![Table 1—Allele and genotype frequency of selected single nucleotide polymorphism (SNP) markers in 96 beef cattle (MARC beef cattle diversity panel version 2.1; MBCDP2.1) and 168 cattle randomly selected at the time of slaughter at a facility that primarily processes culled dairy cattle.](image-url)

* Bos taurus chromosome designation. † The nucleotide alleles for the sense strand are shown in the order of their prevalence in the beef cattle (ie, allele 1 corresponds to the major allele). The allele and genotype frequencies are presented as the fraction of the number of animals genotyped. Frequencies that sum to 0.99 or 1.01 are the result of rounding errors. P$_I$ is the probability of identity and is the average probability that 2 unrelated individuals selected at random would have identical SNP genotypes. Both NOS2 SNPs (AH13-1 and AH13-4, 161 nucleotides apart) were informative, and the haplotype frequencies were used to estimate the P$_I$. NA = Not applicable.
individuals would have the same genotype was expressed as follows: \( P_I = (X_{11})^2 + (X_{12})^2 + (X_{22})^2 \), where \( X_{11} \), \( X_{12} \), and \( X_{22} \) were the relative genotype frequencies of \( A_1A_1 \), \( A_1A_2 \), and \( A_2A_2 \), respectively, in the target population. The \( P_I \) for a combination of multiple SNP markers is the product of the \( P_I \) for each individual marker. The underlying assumption is that the marker spacing is sufficient for meiotic recombination to cause alleles to be randomly associated with one another.

The SNP genotypes were compared between samples by first creating a genotype profile of concatenated SNP genotypes for each sample. In some instances, assay failure resulted in missing genotypes for 1 or more SNP markers in an animal’s profile. In these cases, the SNP genotype was identified as “O” to track missing data in the genotype profile. Concatenated genotype profiles were compared between sample pairs with a Perl script that compared the SNP genotype at each position in profile A with the same position in profile B. This process was repeated for every possible sample pairwise combination of genotype profiles. Numbers of matches (concordant genotypes), mismatches (discordant genotypes), and incomplete genotype profiles were compared between sample pairs with a pairwise fashion (n = 28,224 comparisons [ie, 1682]). The mean number of SNP genotypes available for comparison between 2 genotype profiles was 17.6 (ie, 20 possible SNP genotypes for each sample and a 93.8% genotype scoring rate). When all 28,224 possible pairwise comparisons of blood and liver sample genotypes were examined, the number of discordant SNP genotypes per blood-liver sample pair was normally distributed with mean, mode, and SD of 10, 10, and 2.4, respectively (Figure 1). However, when only the 168 purportedly matched blood-liver sample pairs were examined, the distribution of discordant SNP genotypes appeared bimodal. There were 128 blood-liver sample pairs with \( \leq 2 \) mismatches, and 16 blood-liver sample pairs with \( \geq 6 \) mismatches (mean, 9.6 [52%] mismatches; range, 6 to 14 [33% to 76%] mismatches).

The discordance rate that could be attributed to genotyping error, estimated from results for the 48 split samples, was estimated to be < 1% (8 discordant genotype pairs among 843 evaluated).

Possible causes of mismatching—Direct examination of genotype profiles for all samples revealed 5 samples that did not match their purported sample pair but had a substantial number of genotype matches with another sample. Four of these mismatched cases were the apparent result of a position shift in the sample collection process at the slaughter plant (ie, the genotype of the blood sample from the first animal matched the genotype for the liver sample from the next animal in line). In the fifth case, the genotype profile from the liver sample (purported to be from animal No. 60) matched the genotype profiles from both the blood and the liver sample from animal No. 73, which were collected on a different day. However, when the samples from this case were regenotyped, there were 11 discordant genotypes between samples from animals 60 and 73, demonstrating that the genotype profiles were consistent with their original purported pairs. The liver sample identification numbers for these 2 animals were similar (ie, 20359590 and 20357590), suggesting that a DNA sample label was transposed at some point in the process after collection.

Discussion

Results of the present report suggest that this selected set of 20 bovine SNP markers was sufficiently informative to measure the accuracy of sample tracking in a slaughter plant that processes about 90% dairy cattle.
tles. Moreover, 12 of the 20 SNP markers met previously published criteria for high informativity, whereby the frequency of the minor homozygous genotype was $> 0.10$ (minor allele frequency $> 0.31$; $P_I < 0.416$). Importantly, this genotype frequency cutoff lends significant power in parentage testing situations.

These 20 SNP markers have been shown to be polymorphic in diverse bovine populations and thus may facilitate DNA-based traceback programs for US beef and dairy cattle. Subsequent to the present study, the discovery of BSE in a Holstein cow at a Washington state slaughter facility in December 2003 highlighted the need for accurate and rapid traceability. The SNP markers described in the present report were successfully used to verify pedigree relationships of the Washington state BSE index case and thereby validate the accuracy of physical records derived from the case. Three of the investigators (MPH, MLC, WWL) were able to use the DNA evidence to confirm that the BSE-affected cow was of Canadian origin.

The USDA's FSIS requires that carcasses and their corresponding parts be identified "as being derived from the particular [live] animal" (9 CFR §310.2). This regulation also requires that live-animal identification devices be retained "in such a way to relate them to the carcass." However, we are not aware of any published studies by the USDA or the US FDA evaluating the accuracy of slaughter plant identification systems. Epidemiological traceback by the US FDA and the USDA's Animal and Plant Health Inspection Service is critical for enhancing food safety and controlling animal disease. The FDA uses live-animal identification to trace animals back to their point of origin when drug residues are found in slaughtered samples. The Animal and Plant Health Inspection Service is responsible for the brucellosis and tuberculosis eradication programs (9 CFR, §78 and §77, respectively) and uses live-animal identification and slaughter samples to trace infected animals to their premise of origin. Animal and Plant Health Inspection Service regulation 9 CFR, §78.1 anticipates a large percentage of traceback failure, in part, because of misidentification at slaughter. In states with a class B or C brucellosis designation, Animal and Plant Health Inspection Service regulation requires only "80 percent of all brucellosis reactors be traced to the farm of origin." For states designated as class A or brucellosis free, the regulation requires that at least 90% of all brucellosis reactors be traced to the farm of origin. Although the physical labeling in the present study did not demonstrate a high-fidelity system for maintaining correct animal identification at slaughter, it was within the level of expectations. The actual success of traceback is expected to be higher than the sample matching rate at slaughter because additional information is available in a traceback situation (eg, manual lot and position recordings and carcass weights).

Two important technologic factors that influence the power to discern matched samples are the genotype scoring rate and the genotype scoring error rate. For example, the genotype scoring rate in the present study was 93.8%. Thus, the probability that a particular SNP was scored in both samples was 0.938 $\times$ 0.938. In other words, only 89% of potential SNP genotype comparisons were evaluated because of missing data. On the other hand, genotyping error caused ambiguity in otherwise perfectly matched samples. Thus, when the total number of SNP markers that are used is kept low to gain economy, it is imperative that the genotype scoring rate is high (eg, $> 98\%$) and that the genotype scoring error rate is low (eg, $< 1\%$). One way to overcome these technical barriers is to increase the number of SNP markers used. The combined power of identification for multiple unlinked loci increases geometrically with each informative unlinked locus. Larger, more powerful sets of DNA markers will become available as additional SNP markers are identified for use in US beef and dairy cattle and the price of genotyping decreases.

A variety of factors may have contributed to the sample misidentification rate (9.5%) observed in the present study. For example, the position of some liv ers in the processing line appeared to have shifted with respect to their expected position. Because samples were often collected in groups of 1 or 2, a shift in line position may have resulted in collection of samples with no other genetic matches in the data set. Of the 32 blood and liver samples that did not match their expected sample partners, 27 had no other matches in the data set. Another important factor is sample handling after collection. To generate a genotype, the DNA must typically be extracted, diluted, and transferred to reaction mixtures. Depending on the genotyping method, there may be several biochemical steps required before the genotype is assayed. Each step in this process provides an opportunity for cross-contamination or mislabeling. Although DNA testing is not infallible, it provides a powerful means of verifying the fidelity of animal identification and sample tracking in US beef and dairy cattle.

### References