Bond University
ePublications@bond

Faculty of Health Sciences & Medicine Publications | Faculty of Health Sciences & Medicine

4-5-2008

Forensically relevant SNP classes

Bruce Budowle
FBI Laboratory

Angela van Daal
Bond University, Angela_Van_Daal@bond.edu.au

Follow this and additional works at: http://epublications.bond.edu.au/hsm_pubs

Part of the Genetic Processes Commons, and the Medical Genetics Commons

Recommended Citation
Bruce Budowle and Angela van Daal. (2008) "Forensically relevant SNP classes" *BioTechniques: The international journal of life science methods,* 44 (5), 603-610.

http://epublications.bond.edu.au/hsm_pubs/213

This Journal Article is brought to you by the Faculty of Health Sciences & Medicine at ePublications@bond. It has been accepted for inclusion in Faculty of Health Sciences & Medicine Publications by an authorized administrator of ePublications@bond. For more information, please contact Bond University's Repository Coordinator.

Forensically relevant SNP classes

Bruce Budowle[1] and Angela van Daal[2]

BioTechniques 44:603-610 (25th Anniversary Issue, April 2008)
doi 10.2144/000112806

Forensic samples that contain too little template DNA or are too degraded require alternate genetic marker analyses or approaches to what is currently used for routine casework. Single nucleotide polymorphisms (SNPs) offer promise to support forensic DNA analyses because of an abundance of potential markers, amenability to automation, and potential reduction in required fragment length to only 60-80 bp. The SNP markers will serve an important role in analyzing challenging forensic samples, such as those that are very degraded, for augmenting the power of kinship analyses and family reconstructions for missing persons and unidentified human remains, as well as for providing investigative lead value in some cases without a suspect (and no genetic profile match in CODIS). The SNPs for forensic analyses can be divided into four categories: identity-testing SNPs; lineage informative SNPs; ancestry informative SNPs; and phenotype informative SNPs. In addition to discussing the applications of these different types of SNPs, this article provides some discussion on privacy issues so that society and policymakers can be more informed.

INTRODUCTION

Genetic typing is a powerful means of establishing identity in criminal cases where biological evidence is found at crime scenes—paternity testing, inheritance matters, identification of victims in mass disasters, and identification of missing persons from human remains. Genetic evidence can be derived from any biological material, including blood, semen, bone, hair, teeth, muscle tissue, and saliva, and can be used for characterization of animals, plants, and microorganisms. The repertoire of genetic markers used for characterization of biological materials has evolved substantially within the forensic field with each marker set and concomitant technology platform, augmenting resolution and/or sensitivity of detection. In this paper we briefly describe the current markers used for forensic identification of humans and also what new genetic markers are likely to be developed, namely single nucleotide polymorphisms (SNPs). SNPs offer further opportunities for the forensic scientist to genetically characterize an evidentiary sample or identify an unknown person.

History of Forensic Genetic Typing

Twenty-five years ago polymorphic protein genetic markers were used to potentially differentiate individuals (1–3; Figure 1). However, several factors limited the forensic use of protein-based genetic systems. The discrimination power of these systems is low, so individualization is not possible. Also the proteins are not present at sufficient levels for typing in most tissues, and they are relatively unstable in biological samples exposed to environmental insults.

Typing genetic polymorphisms at the DNA level helps to overcome these limitations to a much greater degree. First, there is a tremendous amount of variation at the DNA level to exploit for identity testing. Second, any biological material that contains nucleated cells potentially can be typed for DNA polymorphisms. Third, DNA tends to be more stable in forensic samples than proteins. Given the current DNA typing techniques and the battery of available genetic markers, typing of human polymorphisms at the DNA level is more sensitive, more specific, and more informative than the classical protein genetic markers. Moreover, DNA technology affords the forensic scientist the greatest potential to exclude individuals who have been falsely associated with a biological sample and to reduce the number of individuals potentially included as contributors of the sample to a few (if not one) individuals.

The first method used routinely for human identity testing (from the mid-1980s) in the United States and many other countries was restriction fragment length polymorphism (RFLP) typing of variable number of tandem repeat (VNTR) loci (4–6). Typing VNTR loci by RFLP analysis provided a very high power of discrimination, but required samples containing at least 10–25 ng of DNA that needed to be relatively intact with fragment lengths up to 10,000 bp for successful typing (7). RFLP analysis was replaced more than a decade ago by amplification of DNA loci by the polymerase chain reaction (PCR) (8) with subsequent typing of genetic markers (Figure 1). PCR enables the analysis of samples containing picogram levels of DNA, thus increasing the sensitivity of detection one to two orders of magnitude. In fact, minute quantities of DNA extracted from many types of samples are typed routinely and successfully using PCR-based assays in forensic laboratories. Sample materials include blood, semen, saliva, and sweat deposited on various substrates including clothing, cigarettes, postage stamps, envelope flaps, drinking straws and containers, chewing gum, and face masks; vaginal swabs from a rape victim; various tissues from human remains; and possible reference samples from personal items, such as hair brushes, toothbrushes, and razors, which may be useful in the identification of unknown remains.

The first genetic marker systems analyzed using PCR-based systems were based on SNP variation. Sequence polymorphisms at the HLA-DQA1 locus and polymarker loci (LDLR, GYPA, HBGG, D7S8, and Gc) were typed by use of allele-specific oligonucleotide (ASO) hybridization probes in a reverse dot blot format (9–12), with a different ASO probe for each allele detected at each

[1]FBI Laboratory, Quantico, VA, USA and [2]Faculty of Health Science & Medicine, Bond University, Gold Coast, Australia

BTN25-RV-BudowleR5.indd 603 4/10/08 2:28:24 PM

locus. The battery of ASO probes was bound to a nylon membrane strip and the PCR-amplified alleles were hybridized to the immobilized probes to which they are complementary. Because an identifier molecule (or tag) was attached to the 5′ end of one of the primers, a detectable label was incorporated into the amplified allelic products. These SNP-based systems offered high sensitivity of detection but did not provide the power of discrimination that VNTR/RFLP typing afforded, and were not very useful for deconvoluting the contributions of mixed samples because of their limited polymorphism.

To increase the power of discrimination of PCR-based systems, VNTR loci were amplified by PCR, with the allelic forms separated by electrophoresis, and subsequently detected by silver staining (13). Short tandem repeat (STR) or microsatellite loci, part of a subclass of the highly polymorphic VNTR loci, have replaced the early PCR-based genetic markers. They now are used in forensic laboratories worldwide (14–16) because of their small product size, polymorphic nature, and semi-automated analytical methods. These repeat sequence loci are abundant in the human genome and are highly polymorphic. The forensically important autosomal STR loci are composed of tandemly repeated sequences, each of which is only four bp in length (with a couple of loci being composed of pentanucleotide repeat units). The number of alleles of the forensically relevant STR loci ranges usually from 5 to 20 common alleles. The regions containing the repeats are generally quite small and are therefore amenable to amplification by PCR. Because the product length is much smaller (generally less than 350 bp) than the length of a DNA fragment required for VNTR analysis by RFLP, somewhat degraded samples are now typeable (17,18). To reduce sample consumption and minimize laboratory manipulations, the STR loci are analyzed in a multiplex fashion, with as many as 15 STR loci amplified and typed simultaneously (19,20). Therefore, the currently used multiplex autosomal STR loci afford high sensitivity, high specificity, high discrimination power, and the ability to type limited quantity or degraded samples, all in a semi-automated manner.

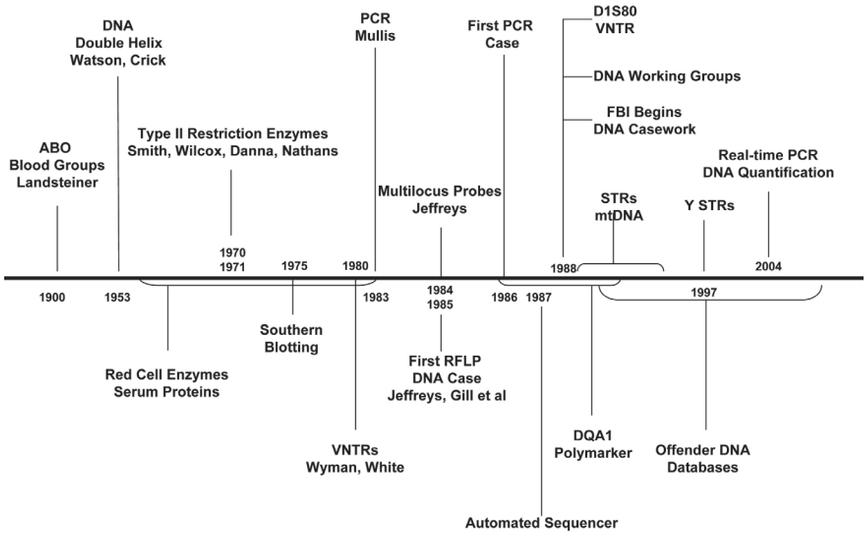

Figure 1. Timeline of selected significant developments impacting forensic genetic analysis and introduction of various genetic marker systems.

SNPs for Human Identification

Even with STR typing there are forensic biological samples that are not amenable to analysis; typically these samples contain too little template DNA or are too degraded. Alternate technologies have been sought to extend further the capability of typing difficult samples by primarily focusing on sequencing of the hypervariable regions of the control region, and to a lesser extent the coding region, of the human mitochondrial genome (21,22). Mitochondrial DNA (mtDNA) typing is carried out by sequencing, which is laborious, time consuming, and costly for the operational laboratory. Therefore, only a portion of forensic DNA laboratories have implemented mtDNA analysis. In addition, the power of discrimination achieved using mtDNA typing is not as high as that provided by the battery of STR loci.

Another class of genetic markers that may prove useful, particularly for typing degraded samples and for increasing the amount of genetic information gleaned from challenged forensic samples, is SNPs (see References 23 and 24 and references within). SNPs are base substitutions, insertions, or deletions that occur at single positions in the genome of any organism. Most SNPs are biallelic and thus are not as informative on a per-locus basis as the forensically selected STR loci. Because all forensic DNA indices, particularly convicted felon DNA databases, are well-established and based on STR loci, it is unlikely that SNPs will become the primary forensic markers. It would take a substantial reduction in cost, increased throughput, and enhanced capabilities to resolve mixed samples for the forensic community to replace the STR loci. Therefore, for the foreseeable future, it is unlikely that SNPs will replace STR loci as the predominate genetic markers for human forensic identification (25). Nevertheless, there is a wealth of genetic information that can be tapped, since approximately 85% of human variation is derived from SNPs (26–28).

SNPs, however, offer advantages for forensic analyses in some situations, such as use in mass disasters and missing persons cases where the DNA may be substantially fragmented. If the resultant degraded fragments are smaller than the required length for STR typing, then no result will be obtained. It is possible to design the PCR for SNP amplification to reduce the required fragment length to only 60–80 bp (23,24,29). Therefore, in highly degraded biological materials, SNP testing should provide more results than STR typing. Another advantage is that substantial research and development currently is underway to improve analytical capabilities, possibly making large multiplex assays and complete automation feasible. With automation, rapid typing of many samples with consistent quality control can be achieved; SNP-based technology may enhance current capabilities in this regard. Lastly, SNPs have relatively low mutation rates and thus are stable genetic markers for lineage-based analyses, such as inheritance cases, missing person cases, and situations where no direct reference sample may be available.

BTN25-RV-BudowleR5.indd 604 4/10/08 2:28:27 PM

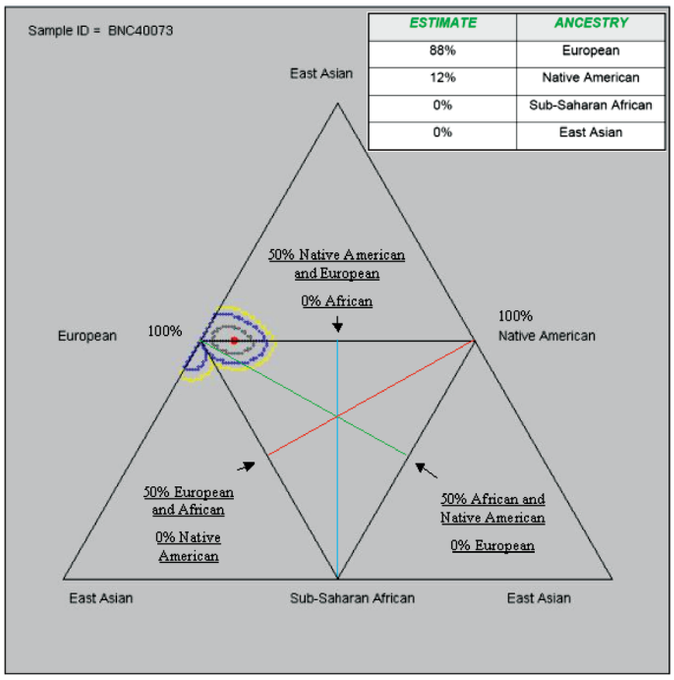

Figure 2. Individual genomic ancestry estimates portrayed with a tetrahedron plot [depicting the four population ancestry informative markers (AIMS) assay]. The most likely estimate (MLE) for an individual is shown with a spot (red), which corresponds to specific percentages shown in the upper righthand box. The percentages are obtained in 4-dimensional space with an algorithm, but are displayed 2-dimensionally; projecting the MLE spot perpendicularly on each of the three axes within any of the four sub-triangles (arrows) gives the corresponding percentage. The closer the spot is to the triangle vertices (labeled European, Sub-Saharan African, East Asian, or Native American), the higher the percentage admixture corresponding to that type of ancestry. The figure is kindly provided by Tony N. Frudakis and reproduced with permission from Reference 40.

SNPs do have a few limitations. Since most SNPs are biallelic, they are less informative for identity testing than STR analyses. Thus, more SNPs will be needed to achieve the same level of discrimination afforded with the core 13 STR loci used in the United States. To obtain the same power of discrimination that existing STR multiplex systems provide, a panel of at least 50–100 SNP autosomal loci would be required (30). Creating such large multiplexes that are validated for forensic analysis is a challenging task. To type such a large battery of SNPs may require far more sample DNA than is needed for several multiplex STR systems. Also SNPs, because of the limited number of alleles per locus, will be more problematic for interpretation in situations involving mixtures, that is, multiple donor samples. Deconvolution of a mixed sample into its individual contributors may be problematic for many scenarios using SNPs. Lastly, the low mutation rate, listed above as an advantage, will contribute to population substructure effects requiring more judicious selection of SNPs for identity-testing purposes.

SNPs for forensic applications should not be discussed generically. Specialized applications are feasible, such as mtDNA and Y chromosome lineage analyses, characterization of highly degraded DNA samples, migration and lineage biographical ancestry analysis, high-throughput rapid screening of DNA samples prior to conventional STR analysis, and potential determination of physical traits. The SNPs for forensic analyses can be divided into four categories:

1. Identity-testing SNPs—for individualization; requiring high heterozygosity and a low coefficient of inbreeding (Fst) (i.e., low population heterogeneity)
2. Lineage informative SNPs—sets of tightly linked SNPs that function as haplotype markers to identify missing persons through kinship analyses
3. Ancestry informative SNPs—for establishing high probability of an individual's biogeographical ancestry to indirectly infer some phenotypic characteristics for investigative lead value; requiring low heterozygosity and high Fst values
4. Phenotype informative SNPs—for establishing high probability that an individual has a particular phenotypic characteristic, such as skin color, hair color, or eye color for investigative lead value

Identity-testing SNPs

Identity-testing SNPs serve the same function as the forensically selected STR loci. They provide genetic information to differentiate people and thus exclude individuals that cannot be the source of an evidentiary sample or cannot be a putative family member. The best SNPs for identity testing are those that have the highest heterozygosity (i.e., 50% heterozygosity for a biallelic system) and low Fst values, because fewer SNPs will be needed to reach high levels of power of discrimination and fewer reference population datasets will be required for statistical assessments for forensic casework. To date, a few attempts to develop identity SNP panels have been reported, such as a 21 SNP panel (31) and a 52 SNP panel (32) by European forensic laboratories. However, the work by Kidd et al. (33) exemplifies the efforts and selection criteria necessary to develop the most useful identity-testing SNP panels. They selected their potential panel based on the population criteria (high heterozygosity and low Fst), screened a large number of publicly available SNPs, and tested a wide variety of sample populations. More than 90,000 candidate SNPs were screened in European Americans, African Americans, and Chinese/Japanese, and 14,638 were found to have an average heterozygosity ≥0.45 in the three populations. These candidate SNPs were further reduced to 2723 SNPs that demonstrated an Fst <0.01 for the three test populations. From this pool 195 SNPs, which were at least 1 Mb apart from one another, were tested in a broader seven-population panel. Again based on population criteria, 35 SNPs were selected for testing in a diverse 40 population sample set. The result of this evaluation was that 19 identity-testing SNPs were identified that have high heterozygosity and low Fst in all 40 populations. Additional studies are being conducted to identify more SNPs that meet the require-

BTN25-RV-BudowleR5.indd 605 4/10/08 2:28:32 PM

ments stated herein for identity testing (34).

Lineage SNPs

Lineage SNPs that reside on the mtDNA genome and the Y chromosome have been identified. Because of a lack of recombination and a low mutation rate, these SNPs have been quite informative for evolutionary studies and kinship analyses, especially in cases where the reference sample(s) and the evidence sample are separated by several generations. To date, a small number of validated SNP assays have been developed for forensic casework. As an example, Coble et al. (35) developed 59 SNPs organized into 8 different multiplex panels that target 18 specific common Caucasian HVI/HVII types to further resolve common mtDNA haplotypes.

The most likely forensic use of lineage SNPs is for missing person cases or mass disaster identifications. Successful identification by genetic testing using kinship analysis is limited by the amount of DNA available for analysis, the number of family members for comparison, and the available genetic markers. The current lineage markers, that is, mtDNA and Y chromosome loci, have limited power of discrimination. However, the HapMap project has identified haploblocks (36–38), which for forensic purposes are SNPs on the autosomes that collectively may serve as lineage-based markers. Although there has been no reported work in this area for forensic applications, it is a promising endeavor that could improve the ability to identify missing persons when the number of family members for comparison is very limited.

Ancestry Informative Markers

In some cases, there is no suspect even after an evidentiary sample STR profile has been searched against an offender DNA database or the information about the true perpetrator may not be reliable, as in some eyewitness accounts. It may be useful for an investigative lead if something about the appearance of the perpetrator were known. The biogeographic ancestry of an individual may indirectly provide some limited information about the general appearance of a person and this could provide some investigative lead value. Forensic STR loci are powerful identity markers, but they are poorly suited for defining the genetic biogeographical ancestry of an individual because of the high degree of allele-sharing among populations. Y chromosome and mtDNA haplotypes are used for reconstructing the evolutionary history of human populations and thus can infer something about the genetic ancestry of an individual. However, uniparental inheritance and limited representation of the human genome do not make these genetic markers good indirect predictors of phenotype. Ancestry informative markers (AIMs) are SNPs distributed throughout the human genome that occur at very different frequencies in different world populations (39–44). They can reveal ancestral origin of a sample/person but do not test for physical characteristics. This indirect method of assessing phenotype is based on the correlation of phenotypic expression with certain elements of human population ancestry structure. For example, skin pigmentation is lighter in Northern Europeans than in any other population; with proper databases defining regional and various global reference ancestral populations with skin pigmentation measurements, light skin color of a genetically defined Northern European donor of an evidence sample can be predicted from a precise genomic ancestry estimate. The genetic ancestry is described in terms of proportional ancestry or admixture. This method requires assessment of the genetic variation that strongly correlates with specific populations and development of databases to quantify the specific AIMs. A commercial kit has been developed by DNAPrint Genomics, Inc. (DNAWitness Bio-Geographical Ancestry Kit; www.dnawitness.net; Sarasota, FL, USA) that uses AIMs to assign ancestry to four major biogeographical groups (Sub-Saharan African, European, East Asian, or Native American) and to determine the relative percentages in cases of admixed peoples (39,44–46) (Figure 2). Two more recent DNAPrint kits have been developed: a 320 AIMs Eurasian panel to elucidate Northern European, Southeastern European, Middle Eastern, and South Asian admixture (EurasianDNA 1.0 product); and a 1476 AIMs European panel to elucidate European subancestry [Southeastern Europe (Armenian, Jewish, Italian, and Greek), Iberian (Spanish, Portuguese), Basque (Spanish/French Pyrenees border), Continental European (German, Irish, English, The Netherlands, French, Swiss, and some Italian), and North Eastern European (Polish, Baltic, Swedish, Norwegian, Finnish, and Russian)] (EuropeanDNA 2.0 product; References 39,47).

This genetic ancestry testing of evidence was successfully applied in the Louisiana Multi-agency Homicide Task Force Investigation in the spring of 2003 involving a Louisiana serial killer. Based on a panel of 73 AIMs (DNAWitness kit, version 2.0), ancestry estimates were presented as proportions with multivariate confidence intervals. The analysis indicated that the source of the evidence sample was primarily African ancestry (85%) with a small amount of Native American admixture (15%). This admixture construct correlated with an individual having a dark complexion (39,40,48) and provided investigative information solely in this regard.

Phenotype Informative SNPs

While AIMs can potentially provide useful information regarding the likely appearance of a suspect, they are indirect measures of the phenotype of an individual that convey only biogeographic ancestry. Thus, they have very limited value for describing the physical appearance of an individual and the informative value must be taken into consideration on a case-by-case basis. DNA markers that describe phenotypic traits would enable a more precise genetic prediction of appearance for investigative leads to identify the perpetrator of a crime. They also may be of value in anthropology studies for the facial reconstruction of unknown human remains (i.e., the skull). A direct method requires an assessment of the genetic variation that strongly affects a specific phenotype as well as development of databases to relate these variants to the specific traits (39). The most obvious descriptors of an individual's appearance are coloring, height, and facial features, which are all highly heritable (49,50). It should therefore be possible to determine the genetic polymorphisms, simple and complex, responsible for these different phenotypic traits.

To date most work on phenotype SNPs has concentrated on pigmentation, since the genetic basis of hair, skin, and eye color is well understood from animal model studies. SNPs in a number of pigmentation genes have been associated with various human hair, skin, and eye color phenotypes. The first human gene shown to be associated with normal pigment variation was the *melanocortin 1 receptor* (*MC1R*) gene. Numerous SNPs in this gene have been linked to red hair and fair skin (51–54). There are three common *MC1R* polymorphisms, R151C, R160W, and D294H, which account for about 90%

BTN25-RV-BudowleR5.indd 606 4/10/08 2:28:34 PM

of red hair and fair skin phenotypes in some populations (55). An R307G variant in the *MC1R* gene has been found in Neanderthals, suggesting that this human species had reduced pigmentation that may have produced fair skin and red hair coloring (56). The *agouti signaling protein* (*ASIP*) gene, encoding an antagonist of the MC1R protein, has also been linked to skin pigmentation (57–59).

A rapid forensic screening assay was developed for 12 *MC1R* variants for the red hair phenotype as a tool to provide investigative leads about the appearance of the donor of an evidentiary sample (60). While the proof of concept was demonstrated, the test itself has limited value. The test panel is highly predictive for homozygotes and compound heterozygotes. Individuals that do not carry the variants typically do not have red hair. Approximately 84% of redheads are detectable with this panel. However, only a small portion of the European population is phenotypically red-haired and even less of non-European-based populations are. More phenotypic information will need to be gleaned per assay to warrant the use as forensic evidence. Multiplexing of a number of physical characteristic SNPs should be a goal for assay development.

Two solute carriers, *SLC45A2* (also known as *MATP)* and *SLC24A5*, are key genes in the determination of human pigmentation phenotype. A third solute carrier gene, *SLC24A4*, has recently been associated with hair and eye color (PA1) (61). The F374L polymorphism in the *SLC45A2* gene shows highly statistically significant associations with dark hair, skin, and eye pigmentation (62). Other SNPs in the promoter region of this gene, which are not in linkage disequilibrium with the F374L polymorphism, are significantly associated with olive skin color (63). The *SLC24A5* gene has a coding polymorphism with an evolutionarily conserved allele predominant in African and East Asian populations. It also has a variant allele that is nearly fixed in European populations that correlates with lighter skin pigmentation (64). Association of the *Kit ligand* (*KITLG*) gene with skin pigmentation was found by searching candidate pigmentation genes for evidence of recent population-specific selective sweeps (65). In a striking example of adaptive evolution Miller et al. (66) show that regulatory changes in this gene contribute to pigment variation in both the stickleback fish and humans based on similar genetic mechanisms. One of the key determinants of human iris pigmentation appears to be the *P* (*OCA2*) gene (39, 67, 68), although alleles of other pigmentation genes such as *tyrosinase-related protein 1* (*TYRP1*) and *dopachrome tautomerase* (*DCT*) also have been associated with eye color (69). Frudakis et al. have shown 33 *OCA2* SNPs in six haploblocks to be highly associated with digitally quantified iris color (70). Recently Eiberg et al. (71) identified two SNPs in a highly conserved intronic region of the gene *HERC2*, which lies upstream of the *P* gene. These SNPs appear to be perfectly associated with blue and brown eye color in a Danish population sample. Luciferase assays were used to demonstrate that this conserved intronic region significantly reduces the activity of the *P* gene promoter. In addition, mobility shift assays showed that the two alleles of one of these SNPs bind different subsets of nuclear extracts. The finding of the same haplotype in the blue-eyed Danish population and a small number of blue-eyed Turkish and Jordanian individuals suggests that there may be a common founder mutation.

While the understanding of the genetics of pigmentation is not complete, it appears that there is a relatively small number of genes that contribute substantially to phenotypic variance. Polymorphisms in three genes, *SLC24A5, TYR,* and *SLC45A2*, have been demonstrated to have highly significant additive associations with skin color that collectively account for a large fraction of the natural variation of skin pigmentation in at least one population group (72). Two genes, *MC1R* and *P*, account for a large portion of the phenotypic variance of red hair color and blue eye color, respectively. The success to date in unraveling much of the genetic basis of pigmentation—a complex, polygenic trait—offers hope that the causative SNPs of other highly heritable traits such as height and facial morphology will be achievable.

Privacy Issues

The use of genetic markers raises concerns about privacy and human rights, and scientists, as informed individuals, should take an active role in protecting such rights. Scientists should provide realistic information about privacy concerns so that society and policymakers can make the most informed decisions about the use of genetic markers and potential concerns (or lack thereof) regarding the violation of the privacy of individuals.

As an example, the identity-testing SNPs, like STRs, provide little or no direct personal or private information about the donor. Thus, it would seem that these loci would not be a concern regarding privacy, and could therefore be endorsed for identity testing. For their SNP panel, Sanchez et al. (32) pragmatically selected SNPs that were located at least 100 kb from any known gene, presumably to minimize potential for perceived privacy risks. When identity-testing SNPs are selected primarily based on the criteria of high heterozygosity and low Fst values, they will not convey significant information about the gene variants for any Mendelian disorder, even if there is complete linkage disequilibrium. Multigenic disorders are less likely to be disclosed in connection with identity-testing SNPs.

To appreciate the impact on privacy for identity-testing SNPs, consider a fictitious dominant Disease Gene (DG) that is rare, such that about 10,000 people in the United States are afflicted with the condition and they all have a common ancestral mutant allele C for DG. Next, assume that 10% of these individuals are in a tested population of 5 million people. If the SNP is biallelic and has a high heterozygosity of 0.50, assuming Hardy-Weinberg expectations, approximately 3,750,000 people in the population will carry at least one copy of the SNP DG C allele (0.25 CC, 0.50 CT). Of these 3,750,000 positive typed individuals, only 1000 will have the condition. Thus, the probability of predicting the DG mutation based on the DG C allele is only 0.00027 (1000/3,750,000). The SNP locus has essentially no predictive value in the population. There is no scientific basis for restricting selection of identity-testing SNPs that have high heterozygosity and low Fst values based on linkage to any gene.

The same privacy issues apply to lineage-based autosomal SNPs. However, in contrast to biallelic SNPs, tightly linked SNP haplotypes will have more than two types. Still, SNP candidates will be sought that yield commonly occurring haplotypes. Thus, they will approximate the value equivalent to a STR locus. Privacy concerns for linkage of a haplotype with, for example, a disease gene, should be relatively low and similar to concerns for the currently used forensic STR loci. If a haplotype were very rare, this issue could be revisited prior to its adoption for forensic applications.

The AIMs test gained notoriety for its use in a Louisiana case where the genetic evidence convinced the police to abandon their dragnet on 700 Caucasian men. Instead of using less reliable subjective

BTN25-RV-BudowleR5.indd 607 4/10/08 2:28:36 PM

eyewitness accounts for generating investigative leads, investigators refocused their efforts, broadening their search because of the ancestry results provided by a panel of AIMs (39,40,48). Although not precisely predictive, AIMs can offer objective, repeatable, and quantifiable scientific evidence resulting in a physical, although limited, profile. The resulting genetic profile is probabilistic and population definitions are not always clearly delineated. It should be recognized that some admixture proportions will be better correlated with phenotype than others; thus some cases will yield better investigative lead value than other cases. However, the possible linkage to a disease gene might be a criterion to consider regarding privacy. Lastly, there may be a potential prejudice that can result from the prediction of ancestry. Balance must be struck between the usefulness of the predicted physical appearance based indirectly on ancestry versus potential for racial and ethnic discrimination. While this needs to be addressed by society and policymakers, the eventual ability to directly predict physical appearance may replace the use of AIMs.

Prediction of physical appearance (limited to skin, eye, and hair color and possibly height and facial features) is less likely to be considered a privacy issue. Given that a person's appearance is readily discernible without DNA typing, the use of DNA polymorphisms to predict phenotype should not be controversial. Determining physical appearance of the donor of an evidence sample based on empirical scientific techniques can produce objective, quantitative, and testable data. In addition, phenotypic SNP testing would only be used for anonymous crime scene samples for investigative purposes. On arrest, the STR profile of the suspect, not the phenotypic SNP profile, would be generated and compared with the evidence profile (provided there was sufficient DNA available).

CONCLUSION

SNPs are unlikely to replace the battery of STR loci in the foreseeable future. However, SNP markers will serve an important role in analyzing challenging forensic samples, such as those that are very degraded, for augmenting the power of kinship analyses and family reconstructions for missing persons and unidentified human remains, as well as for providing investigative lead value in some cases without a suspect (and no STR match in CODIS). There are four different classes of SNPs that apply to forensic analyses: identity-testing SNPs, lineage informative SNPs, ancestry informative SNPs, and phenotypic SNPs. The identity-testing SNPs are likely the first group to be implemented for routine work, with lineage based systems to follow. The ancestry SNPs (which already have been used in some cases) and phenotypic SNPs will be used in more specialized applications that may provide useful investigative leads. While more accurate than AIMs for predicting phenotype, phenotypic SNPs require considerably more research and development efforts and to date have been most applicable to hair and eye color. We anticipate that more efforts will be dedicated to the development of SNPs to enhance the forensic science toolbox for genetically characterizing biological evidence with the most challenging of forensic samples while generating valid, objective, and reliable results in a timely fashion. Lastly, we have provided some discussion on privacy issues so that society and policymakers can be more informed when deciding on the selection criteria for genetic markers and their impact on the privacy of an individual.

ACKNOWLEDGMENTS

This is publication number 08-03 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation. We would like to thank Tony Frudakis (DNAPrint Genomics, Inc., Sarasota, FL, USA) and Ron Fourney (Royal Canadian Mounted Police) for their contributions and thoughtful comments.

COMPETING INTERESTS STATEMENT

The authors declare no competing interests.

REFERENCES

1. **Sensabaugh, G.F.** 1995. Biochemical markers of individuality p. 338-415. *In* R. Saferstein (Ed.), Criminalistics—An Introduction to Forensic Science, 5th ed. Prentice Hall Education, Upper Saddle River, NJ.
2. **Budowle, B. and R.C. Allen.** 1987. Electrophoresis reliability. I. The contaminant issue. J. Forensic Sci. *32*:1537-1550.
3. **Murch, R.S. and B. Budowle.** 1986. Applications of isoelectric focusing in forensic serology. J. Forensic Sci. *31*:869-880.
4. **Budowle, B. and F.S. Baechtel.** 1990. Modifications to improve the effectiveness of restriction fragment length polymorphism typing. Appl. Theor. Electrophor. *1*:181-187.
5. **Jeffreys, A.J., V. Wilson, and S.L. Thein.** 1985. Hypervariable minisatellite regions in human DNA. Nature *314*:67-73.
6. **Wyman, A.R. and R. White.** 1980. A highly polymorphic locus in human DNA. Proc. Natl. Acad. Sci. USA *77*:6754-6758.
7. **Giusti, A.M. and B. Budowle.** 1995. Chemiluminescence-based detection system for human DNA quantitation and restriction fragment length polymorphism (RFLP) analysis. Appl. Theor. Electrophor. *5*:89-98.
8. **Saiki, R.K., S. Scharf, F. Faloona, K.B. Mullis, G.T. Horn, H.A. Erlich, and N. Arnheim.** 1985. Enzymatic amplification of beta-globin genomic sequences and restriction analysis for diagnosis of sickle cell anemia. Science *230*:1350-1354.
9. **Budowle, B., J.A. Lindsey, J.A. DeCou, B.W. Koons, A.M. Giusti, and C.T. Comey.** 1995. Validation and population studies of the loci LDLR, GYPA, HBGG, D7S8, and Gc (PM loci), and HLA-DQα using a multiplex amplification and typing procedure. J. Forensic Sci. *40*:45-54.
10. **Comey, C.T. and B. Budowle.** 1991. Validation studies on the analysis of the HLA-DQ alpha locus using the polymerase chain reaction. J. Forensic Sci. *36*:1633-1648.
11. **Gyllensten, U.B. and H.A. Erlich.** 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQ alpha locus. Proc. Natl. Acad. Sci. USA *85*:7652-7656.
12. **Walsh, P.S., N. Fildes, A.S. Louie, and R. Higuchi.** 1991. Report of the blind trial of the Cetus AmpliType HLA DQα forensic deoxyribonucleic acid (DNA) amplification and typing kit. J. Forensic Sci. *36*:1551-1556.
13. **Budowle, B., R. Chakraborty, A.M. Giusti, A.J. Eisenberg, and R.C. Allen.** 1991. Analysis of the variable number of tandem repeats locus D1S80 by the polymerase chain reaction followed by high resolution polyacrylamide gel electrophoresis. Am. J. Hum. Genet. *48*:137-144.
14. **Edwards, A., A. Civitello, H.A. Hammond, and C.T. Caskey.** 1991. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am. J. Hum. Genet. *49*:746-756.
15. **Budowle, B., T.R. Moretti, S.J. Niezgoda, and B.L. Brown.** 1998. CODIS and PCR-based short tandem repeat loci: law enforcement tools, p. 73–88. *In* Second European Symposium on Human Identification 1998, Promega Corporation, Madison, WI.
16. **Budowle, B., B. Shea, S. Niezgoda, and R. Chakraborty.** 2001. CODIS STR loci data from 41 sample populations. J. Forensic Sci. *46*:453-489.
17. **Whitaker, J.P., T.M. Clayton, A.J. Urquhart, E.S. Millican, T.J. Downes, C.P. Kimpton, and P. Gill.** 1995. Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. BioTechniques *18*:670-677.
18. **Ballantyne, J.** 1997. Mass disaster genetics. Nat. Genet. *15*:329-331.
19. **Collins, P.J., L.K. Hennessy, C.S. Leibelt, R.K. Roby, D.J. Reeder, and P.A. Foxall.** 2004. Developmental validation of a single-tube amplification of the 13 CODIS loci, D2S1338, D19S433, and amelogenin: the AmpFlSTR® Identifiler® PCR amplification kit. J. Forensic Sci. *49*:1265-1277.
20. **Krenke, B.E., A. Tereba, S.J. Anderson, E. Buel, S. Culhane, C.J. Finis, C.S. Tomsey, J.M. Zachetti, et al.** 2002. Validation of a 16-locus fluorescent multiplex system. J. Forensic Sci. *47*:773-785.
21. **Wilson, M.R., J.A. DiZinno, D. Polansky, J. Replogle, and B. Budowle.** 1995. Validation of mi-

BTN25-RV-BudowleR5.indd 608 4/10/08 2:28:39 PM

tochondrial DNA sequencing for forensic casework analysis. Int. J. Legal Med. *108*:68-74.
22. **Budowle, B., M.W. Allard, M.R. Wilson, and R. Chakraborty.** 2003. Forensics and mitochondrial DNA: applications, debates, and foundations. Annu. Rev. Genomics Hum. Genet. *4*:119-141.
23. **Budowle, B.** 2004. SNP typing strategies. Forensic Sci. Int. *146*:S139-S142.
24. **Budowle, B., J. Planz, R. Campbell, and A. Eisenberg.** 2004. SNPs and microarray technology in forensic genetics: development and application to mitochondrial DNA. Forens. Sci. Rev. *16*:22-36.
25. **Gill, P., D.J. Werrett, B. Budowle, and R. Guerrieri.** 2004. An assessment of whether SNPs will replace STRs in national DNA databases—joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). Sci. Justice *44*:51-53.
26. **Cooper, D.N., B.A. Smith, H.J. Cooke, S. Niemann, and J. Schmidtke.** 1985. An estimate of unique DNA sequence heterozygosity in the human genome. Hum. Genet. *69*:201-205.
27. **Holden, A.L.** 2002. The SNP consortium: summary of a private consortium effort to develop an applied map of the human genome. BioTechniques *32*:S22-S26.
28. **Wang, D.G., J.B. Fan, C.J. Siao, A. Berno, P. Young, R. Sapolshy, G. Chandona, N. Perkins, et al.** 1998. Large scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science *280*:1077-1082.
29. **Divne, A.M. and M. Allen.** 2005. A DNA microarray system for forensic SNP analysis. Forens. Sci Int. *154*:111-121.
30. **Chakraborty, R., D.N. Stivers, B. Su, Y. Zhong, and B. Budowle.** 1999. The utility of STR loci beyond human identification: implications for the development of new DNA typing systems. Electrophoresis *20*:1682-1696.
31. **Dixon, L.A., C.M. Murray, E.J. Archer, A.E. Dobbins, P. Koumi, and P. Gill.** 2005. Validation of a 21 locus autosomal SNP multiplex for forensic identification purposes. Forensic Sci. Int. *154*:62-77.
32. **Sanchez, J.J., C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, et al.** 2006. A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis *27*:1713-1724.
33. **Kidd, K.K., A.J. Pakstis, W.C. Speed, E.L. Grigorenko, S.L. Kajuna, N.J. Karoma, S. Kungulilo, J.J. Kim, et al.** 2006. Developing a SNP panel for forensic identification of individuals. Forensic Sci. Int. *164*:20-32.
34. **Pakstis, A.J., W.C. Speed, J.R. Kidd, and K.K. Kidd.** 2007. Candidate SNPs for a universal individual identification panel. Hum. Genet. *121*:305-317.
35. **Coble, M.D., R.S. Just, J.E. O'Callaghan, I.H. Letmanyi, C.T. Peterson, J.A. Irwin, and T.J. Parsons.** 2004. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int. J. Legal Med. *118*:137-146.
36. **Daly, M.J., J.D. Rioux, S.F. Schaffner, T.J. Hudson, and E.S. Lander.** 2001. High-resolution haplotype structure in the human genome. Nat. Genet. *29*:229-232.
37. **Gabriel, S.B., S.F. Schaffner, H. Nguyen, J.M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, et al.** 2002. The structure of haplotype blocks in the human genome. Science *296*:2225-2229.
38. **Tishkoff, S.A. and B.C. Verrelli.** 2003. Role of evolutionary history on haplotype block structure in the human genome: implications for disease mapping. Curr. Opin. Genet. Dev. *13*:569-575.
39. **Frudakis, T. N.** 2008. Molecular Photofitting: Predicting Ancestry and Phenotype from DNA. Academic Press Publishers (Elsevier), Amsterdam, Netherlands.
40. **Frudakis, T.** The inference of phenotype from crime scene. DNA. Forensics Encyclopedia (In press.)
41. **Frudakis, T., K. Venkateswarlu, M.J. Thomas, Z. Gaskin, S. Ginjupalli, S. Gunturi, V. Ponnuswamy, S. Natarajan, and P.K. Nachimuthu.** 2003. A classifier for the SNP-based inference of ancestry. J. Forensic Sci. *48*:771-782.
42. **Rosenberg, N.A., J.K. Pritchard, J.L. Weber, H.M. Cann, K.K. Kidd, L.A. Zhivotovsky, and M.W. Feldman.** 2002. Genetic structure of human populations. Science *298*:2381-2385.
43. **Shriver, M.D., E.J. Parra, S. Dios, C. Bonilla, H. Norton, C. Jovel, C. Pfaff, C. Jones, et al.** 2003. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum. Genet. *112*:387-399.
44. **Shriver, M.D., M.W. Smith, L. Jin, A. Marcini, J.M. Akey, R. Deka, and R.E. Ferrell.** 1997. Ethnic-affiliation estimation by use of population-specific DNA markers. Am. J. Hum. Genet. *60*:957-964.
45. **Pritchard, J.K., M. Stephens, and P. Donnelly.** 2000. Inference of population structure using multilocus genotype data. Genetics *155*:945-959.
46. **Halder, I., M. Shriver, M. Thomas, J. Fernandez, and T. Frudakis.** A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. Hum. Mutat. (In press.)
47. **Bauchet, M., B. McEvoy, L. Pearson, E. Quillen, T. Sarkisian, K. Hovhannesyan, R. Deka, D. Bradley, and M. Shriver.** 2007. Measuring European population stratification with microarray genotype data. Am. J. Hum. Genet. *80*:948-956.
48. **Shriver, M., T. Frudakis, and B. Budowle.** 2005. Getting the science and the ethics right in forensic genetics. Nat. Genet. *37*:449-450.
49. **Silventoinen, K., S. Sammalisto, M. Perola, D.I. Boomsma, B.K. Cornes, C. Davis, L. Dunkel, M. De Lange, et al.** 2003. Heritability of adult body height: a comparative study of twin cohorts in eight countries. Twin Res. *6*:399-408.
50. **Clark, P., A.E. Stark, R.J. Walsh, R. Jardine, and N.G. Martin.** 1981. A twin study of skin reflectance. Ann. Hum. Biol. *8*:529-541.
51. **Flanagan, N., E. Healy, A. Ray, S. Philips, C. Todd, I.J. Jackson, M.A. Birch-Machin, and J.L. Rees.** 2000. Pleiotropic effects of the melanocortin 1 receptor (*MC1R*) gene on human pigmentation. Hum. Mol. Genet. *9*:2531-2537.
52. **Rees, J.L.** 2000. The melanocortin 1 receptor (MC1R): more than just red hair. Pigment Cell Res. *13*:135-140.
53. **Valverde, P., E. Healy, S. Sikkink, F. Haldane, A.J. Thody, A. Carothers, I.J. Jackson, and J.L. Rees.** 1996. The Asp84Glu variant of the melanocortin 1 receptor (*MC1R*) is associated with melanoma. Hum. Mol. Genet. *5*:1663-1666.
54. **Sturm, R.A.** 2006. A golden age of human pigmentation genetics. Trends Genet. *22*:464-468.
55. **Sturm, R.A., D.L. Duffy, N.F. Box, R.A. Newton, A.G. Shepherd, W. Chen, L.H. Marks, J.H. Leonard, and N.G. Martin.** 2003. Genetic association and cellular function of MC1R variant alleles in human pigmentation. Ann. N. Y. Acad. Sci. *994*:348-358.
56. **Lalueza-Fox, C., H. Römpler, D. Caramelli, C. Stäubert, G. Catalano, D. Hughes, N. Rohland, E. Pilli, et al.** 2007. A melanocortin 1 receptor allele suggests varying pigmentation among Neanderthals. Science *318*:1453-1455.
57. **Kanetsky, P.A., J. Swoyer, S. Panossian, R. Holmes, D. Guerry, and T.R. Rebbeck**. 2002. A polymorphism in the agouti signaling protein gene is associated with human pigmentation. Am. J. Hum. Genet. *70*:770-775.
58. **Voisey, J., N.F. Box, and A. van Daal.** 2001. A polymorphism study of the human *agouti* gene and its association with MC1R. Pigment Cell Res. *14*:264-267.
59. **Voisey, J., C. Gomez-Cabrera Mdel, D.J. Smit, J.H. Leonard, R.A. Sturm, and A. van Daal.** 2006. A polymorphism in the agouti signaling protein (ASIP) is associated with decreased levels of mRNA. Pigment Cell Res. *19*:226-231.
60. **Grimes, E.A., P.J. Noake, L. Dixon, and A. Urquhart.** 2001. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. Forensic Sci. Int. *122*:124-129.
61. **Sulem, P., D.F. Gudbjartsson, S.N. Stacey, A. Helgason, T. Rafnar, K.P. Magnusson, A. Manolescu, A. Karason, et al.** 2007. Genetic determinants of hair, eye and skin pigmentation in Europeans. Nat. Genet. *39*:1443-1452.
62. **Graf, J., R. Hodgson, and A. van Daal.** 2005. Single nucleotide polymorphisms in the MATP gene are associated with normal human pigmentation variation. Hum. Mutat. *25*:278-284.
63. **Graf J., J. Voisey, I. Hughes, A. van Daal.** 1007. Promoter polymorphisms in the *MATP* (*SLC45A2*) gene are associated with normal human skin color variation. Hum. Mut. *28*:710-717.
64. **Lamason, R.L., M.A. Mohideen, J.R. Mest, A.C. Wong, H.L. Norton, M.C. Aros, M.J. Jurynec, X. Mao, et al.** 2005. SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science *310*:1782-1786.
65. **Williamson, S.H., M.J. Hubisz, A.G. Clark, B.A. Payseur, C.D. Bustamante, and R. Nielsen.** 2007. Localizing recent adaptive evolution in the human genome. PLoS Genet. *3*:e90.
66. **Miller, C.T., S. Beleza, A.A. Pollen, D. Schluter, R.A. Kittles, M.D. Shriver, and D.M. Kingsley.** 2007. cis-regulatory changes in kit ligand expression and parallel evolution of pigmentation in sticklebacks and humans. Cell *131*:1179-1189.
67. **Duffy, D.L., N.F. Box, W. Chen, J.S. Palmer, G.W. Montgomery, M.R. James, N.K. Hayward, N.G. Martin, and R.A. Sturm.** 2004. Interactive effects of MC1R and OCA2 on melanoma risk phenotypes. Hum. Mol. Genet. *13*:447-461.
68. **Rebbeck, T.R., P.A. Kanetsky, A.H. Walker, R. Holmes, A.C. Halpern, L.M. Schuchter, D.E. Elder, and D. Guerry.** 2002. *P* gene as an inherited biomarker of human eye color. Cancer Epidemiol. Biomarkers Prev. *11*:782-784.
69. **Frudakis, T., M. Thomas, Z. Gaskin, K. Venkateswarlu, K.S. Chandra, S. Ginjupalli, S. Gunturi, S. Natrajan, et al.** 2003. Sequences associated with human iris pigmentation. Genetics *165*:2071-2083.
70. **Frudakis, T., T. Terravainen, and M. Thomas.** 2007. Multilocus OCA2 genotypes specify human iris colors. Hum. Genet. *122*:311-326.
71. **Eiberg, H., J. Troelsen, M. Nielsen, A. Mikkelsen, J. Mengel-From, K.W. Kjaer, and L. Hansen.** 2008. Blue eye color in humans may be caused by a perfectly associated founder mutation in a regulatory element located within the *HERC2* gene inhibiting *OCA2* expression. Hum. Genet. *123*:117-187.
72. **Stokowski, R.P., P.V. Pant, T. Dadd, A. Fereday, D.A. Hinds, C. Jarman, W. Filsell, R.S. Ginger, et al.** 2007. A genomewide association study of skin pigmentation in a South Asian population. Am. J. Hum. Genet. *81*:1119-1132.

Address correspondence to Bruce Budowle, FBI Laboratory, 2501 Investigation Parkway, Quantico, VA 22135, USA. e-mail: Bruce.Budowle@ic.fbi.gov

To purchase reprints of this article, contact: Reprints@BioTechniques.com

BTN25-RV-BudowleR5.indd 610 4/10/08 2:28:43 PM