Molecular Ecology (2004) doi: 10.1111/j.1365-294X.2004.02126.x

Development of microarray-based diagnostics of voles and shrews for use in biodiversity monitoring studies, and evaluation of mitochondrial cytochrome oxidase I vs. cytochrome *b* as genetic markers

M. PFUNDER,* O. HOLZGANG† and J. E. FREY*
*Agroscope FAW Wädenswil, Swiss Federal Research Station for Horticulture, Department of Crop Protection, Laboratory for Molecular Diagnostics, CH-8820 Wädenswil, Switzerland, †Swiss Ornithological Institute, CH-4209 Sempach, Switzerland

Abstract

Molecular methods are widely used for species identification of mammals. In particular, the mitochondrial cytochrome *b* gene sequence has proven helpful for this purpose. Microarray technology can now open up new perspectives for biodiversity monitoring. With microarrays, many thousands of genetically based characteristics can be tested on one microscopic glass slide called a 'chip'. A 'Mammalia-Chip', for example, could include redundant diagnostic markers to unambiguously identify all European mammal species. Of broader use, and therefore economically more relevant, could be a 'Biodiversity-Chip', containing diagnostic features to distinguish key species in the taxa of bacteria, lichen, molluscs, insects, fungi, mammals, etc. An important prerequisite for any mixed-phyla chip is a standardization of methods. One of the most promising genes as a universal marker for all eukaryotes is cytochrome oxidase I. We show that cytochrome oxidase I is adequate for the discrimination of different species of voles and shrews with cluster analysis. Based on these results we present a diagnostic microarray-chip using cytochrome oxidase I sequences for the identification of three species of *Sorex* (Soricidae, Insectivora) and four species of *Microtus* (Arvicolinae, Rodentia). We conclude that cytochrome oxidase I can be used as an alternative marker to cytochrome *b* in a mixed-phyla chip, or both genes can be used in combination to enhance redundance and thus robustness of a specific chip including small mammals.

Keywords: chip, Insectivora, microarray, monitoring, Rodentia, small mammals

Received 1 October 2003; revision received 19 December 2003; accepted 19 December 2003

Introduction

Many molecular studies have been performed to unravel the relationship of closely related species and to describe newly detected species of mice (Murinae, Rodentia), voles (Arvicolinae, Rodentia) and shrews (Soricidae, Insectivora). The methods include karyotyping (e.g. Mazurok *et al*. 2001; Brünner *et al*. 2002), protein electrophoresis (e.g. Vogel *et al*. 1991) and mitochondrial DNA (mtDNA) analysis (e.g. Fumagalli *et al*. 1999; Martin *et al*. 2000; Mazurok *et al*. 2001; Michaux *et al*. 2001, 2002; Stewart *et al*. 2002).

Biodiversity studies of small mammals, however, are generally based on life trapping and identification by phenotypic characters (Gurnell & Flowerdew 1994; Wilson *et al*. 1996; but also see Riddle *et al*. 2003). Some species of mice, voles and shrews can easily be distinguished by eye, and trapped animals can generally soon be released. However, especially in regions where sibling species coexist, the collection of dead specimens is often necessary. For example, the red-toothed shrews *Sorex araneus*, *S. coronatus* and *S. antinorii* coexist in Switzerland and exhibit rather uniform external features. They can be differentiated only by cranial morphology, chromosome number, protein electrophoresis or molecular analysis (Churchfield 1990; Hausser 1995; Brünner *et al*. 2002). The differentiation of different white-toothed shrew species, such as the three locally coexisting *Crocidura russula*, *C. suaveolens* and *C. leucodon* needs exceptional experience (Mitchell-Jones *et al*. 1999). The subfamily Arvicolinae is also represented by

Correspondence: Dr Monika Pfunder. Tel.: +41 1 783 6187; Fax: +41 1 783 6434; E-mail: monika.pfunder@faw.admin.ch

closely related and similar looking species such as *Microtus agrestis* and *M. arvalis*, both occurring sympatrically in continental Europe (except Scandinavia). The genus *Apodemus* (subfamily Murinae) shows a high degree of geographical phenotypic variation in Europe, and certain species are difficult to distinguish unambiguously in the field. Examples are the three sibling species *Apodemus flavicollis*, *A. sylvaticus* and *A. alpicola* that coexist in the alpine region (Hausser 1995). All these species can be distinguished unambiguously using mitochondrial cytochrome *b* as a molecular marker (Fumagalli *et al*. 1999; Martin *et al*. 2000; Michaux *et al*. 2001, 2002; Jaarola & Searle 2002; Haynes *et al*. 2003).

With the fast improvements in molecular biology and technology, alternatives to 'classical' DNA sequencing are being developed that present interesting perspectives for biodiversity monitoring studies. Among the most promising techniques is microarray technology which allows the parallel analysis of > 10 000 separate features on a glass slide ~ 2 × 8 cm called a 'chip'. This technology is already being used in biology involving studies on microbial communities (e.g. Bodrossy *et al*. 2003; Stine *et al*. 2003), and diagnostics of different bacteria (e.g. Kingsley *et al*. 2002; Rudi *et al*. 2003) or viral pathogens (e.g. Wang *et al*. 2002; Boonham *et al*. 2003; Bystřická *et al*. 2003). The potential for the use of microarrays is enormous: for example, a 'Mammalia-Chip' could include multiple redundant tests to distinguish all European mammals using species-specific DNA sequences. Economically, however, the development of a chip only pays if one kind of chip, produced in large numbers, can be used for many purposes and to answer many different questions in ecology (Rubenstein 2003). Of wider use might, therefore, be a 'Biodiversity-Chip', holding many thousand diagnostic features to distinguish key species of different taxa such as bacteria, lichen, fungi, plants and animals.

Whatever approach is taken, an important prerequisite for the miniaturization and parallelization of diagnostics is the standardization of molecular methods. Cytochrome oxidase I is one of the most promising genes for use as a universal marker for all eukaryotes. Hebert *et al*. (2003) presented a phylum profile of 100 sequences, in which each sequence was derived from a different family and represented one of the following phyla: Annelida, Arthropoda, Chordata, Echinodermata, Mollusca, Nematoda and Plathelmintes. All species from these phyla showed distinct cytochrome oxidase I sequences. Zardoya & Meyer (1996) tested cytochrome oxidase I and considered it to be as good as cytochrome *b* for the resolution of recently diverged vertebrate species. To date, however, not many mammalian cytochrome oxidase I sequences have been analysed. A GenBank search in February 2003 revealed 34 676 cytochrome *b* sequences. Of these, 10 855 hits (31.3%) belonged to mammalian sequences. Searching for cytochrome oxidase I gave 16 605 hits, with only 528 (3.2%) belonging to mammal species. Of these, 61% (321 hits) belonged to *Rattus norvegicus* alone.

The aim of this study was to test the suitability of cytochrome oxidase I for molecular species identification of closely related small mammals as a complement to cytochrome *b* in microarray technology. A chip, including three shrew species, *Sorex coronatus*, *S. araneus* and *S. minutus* (Soricidae, Insectivora), as well as four vole species, *Microtus arvalis*, *M. agrestis*, *M. subterraneus* and *M. nivalis* (Arvicolinae, Rodentia), was designed to illustrate the principle and usefulness of a microarray approach. This study is linked to a Swiss national long-term monitoring and biodiversity study of Module Z3 (Hintermann *et al*. 2002). Because biodiversity studies often include handling of threatened species and monitoring studies should be repeatable over time, this work was carried out using entirely noninvasive methods: sampling was conducted using life traps and the genetic analysis was based almost entirely on hair samples from trapped animals that were later released.

Materials and methods

Sample collection and extraction

Capturing individuals from the taxa Murinae, Arvicolinae and Soricidae was done using life traps (Longworth, TripTrap and Sherman). Trapping was conducted in Switzerland in 2002 / 2003 during the course of an annual biodiversity monitoring schedule for Module Z3 (Hintermann *et al*. 2002). Fresh hair samples were collected from living animals using a pair of tweezers to pull out three small tufts from each individual, cleaning the tweezers with 70% alcohol between individuals. Hair samples were stored dry in the freezer as soon as possible, generally within 12 h. The collection was supplemented with four donated specimens (Table 1). Single hair tufts of ~10–20 hairs including follicles were used for genomic DNA extraction following the protocol of the Gene Elute Mammalian Kit (Sigma) for tissue extraction. For the few dead specimens, tissue material was sampled (Table 1), following the protocol used for hair samples.

DNA amplification and sequencing

Each specimen was identified on the basis of phenotypic characters during sampling. However, because the animals were released afterwards, and identification was therefore not reproducible, we used sequences of cytochrome *b* to confirm field determination. To obtain primers that could be used to amplify a subunit of cytochrome *b* from rodents as well as insectivores, we checked the regions around the two universal primers L14724 (Irwin *et al*. 1991) and H15573 (Taberlet *et al*. 1992) within an alignment of all published cytochrome *b* sequences of Murinae, Arvicolinae and

Table 1 All specimen sampled for this study were collected in Switzerland, most of them in the canton of Lucerne (LU), five in the canton Obwalden (OW), two in Ticino (TI) and one in Aargau (AG). They were either sampled during the biodiversity monitoring module Z3 (Hintermann *et al*. 2002) or provided by U. Käser (P33), T. Maddalena (P50, P51) and A. Pfunder (P45). DNA extraction was done from hair samples (h) and in few cases from tissue (t). GenBank Accession nos are given for cytochrome *b* (cyt *b*) as well as for cytochrome oxidase I (COI)

GenBank Accession no. cyt *b*	GenBank Accession no. COI	Species (Synonym)	Alias	Locality (Switzerland)	Specimen	DNA extr.
Insectivora, Soricidae						
AY332688	AY332664	*Sorex coronatus*	Sorcor	Napf, LU	P01	t
AY332689	AY332665	*Sorex coronatus*	Sorcor	Napf, LU	P02	t
AY332690	AY332666	*Sorex coronatus*	Sorcor	Gläng, LU	P21	h
AY332691	AY332667	*Sorex coronatus*	Sorcor	Gläng, LU	P35	h
AY332692	AY332668	*Sorex coronatus*	Sorcor	Schötz, LU	P38	h
AY332693	AY332669	*Sorex araneus*	Sorara	Pilatus, OW	P54	h
AY332694	AY332670	*Sorex araneus*	Sorara	Pilatus, OW	P55	h
AY332695	AY332671	*Sorex minutus*	Sormin	Gläng, LU	P23	h
AY332696	AY332672	*Sorex minutus*	Sormin	TI	P51	h
no sequence	AY332673	*Neomys fodiens*	Neofod	Gläng, LU	P29	h
AY332697	AY332674	*Crocidura russula*	Crorus	Schötz, LU	P07	h
AY332698	AY332675	*Crocidura russula*	Crorus	Schötz, LU	P09	h
AY332699	AY332676	*Crocidura russula*	Crorus	Schenkon, LU	P33	h, t
Rodentia, Arvicolinae						
AY332706	AY332678	*Clethrionomys glareolus*	Clegla	Schötz, LU	P17	h
AY332707	AY332679	*Clethrionomys glareolus*	Clegla	Bodenberg, LU	P32	t
AY332708	AY332680	*Arvicola terrestris*	Arvter	Schötz, LU	P20	t
AY332709	AY332681	*Arvicola terrestris*	Arvter	Kriens, LU	P37	h
AY332710	AY332682	*Microtus arvalis*	Micarv	Schötz, LU	P06	h
AY332711	AY332683	*Microtus arvalis*	Micarv	Schötz, LU	P11	h
AY332712	AY332684	*Microtus agrestis*	Micagr	Wauwil, LU	P41	h
AY332713	no sequence	*Microtus* (*Pitymys*) *multiplex*	Micmul	Fusio, TI	P50	h
AY332714	AY332685	*Microtus* (*Pitymys*) *subterraneus*	Micsub	Pilatus, OW	P56	h
AY332715	AY332686	*Microtus* (*Chionomys*) *nivalis*	Micniv	Pilatus, OW	P58	h
AY332716	AY332687	*Microtus* (*Chionomys*) *nivalis*	Micniv	Pilatus, OW	P59	h
Rodentia, Murinae						
AY332700	no sequence	*Apodemus sylvaticus*	Aposyl	Wallbrig, LU	P40	h
AY332701	no sequence	*Apodemus sylvaticus*	Aposyl	Wallbrig, LU	P46	h
AY332702	no sequence	*Apodemus flavicollis*	Apofla	Schötz, LU	P15	h
AY332703	no sequence	*Apodemus flavicollis*	Apofla	Schötz, LU	P39	h
AY332704	no sequence	*Apodemus flavicollis*	Apofla	Badachtal, LU	P47	h
AY332705	AY332677	*Mus musculus domesticus*	Musmus	Thalheim, AG	P45	h

Soricidae by August 2002, in GenBank. The newly designed primers used in this study are forward primer F-muarso (5′-ATGACATGAAAAATCATYGTTGT-3′) and reverse primer R-muarso (5′-GAAATATCATTCKGGTTTAATRTG-3′).

The polymerase chain reaction (PCR) was performed in a total volume of 20 μL containing 1 μL of undiluted extract, 6 μM of each oligonucleotide primer and 10 μL of HotStar Taq™ MasterMix (Qiagen). Amplification was conducted using a ramping protocol on Perkin–Elmer GeneAmp 9600, where an initial denaturation step at 95 °C for 15 min was followed by 45 cycles of 40 s at 94 °C, 15 s at 45 °C, 30 s ramping to 60 °C, 2 min elongation at 72 °C and final extension at 72 °C for 7 min. PCR products and negative controls were checked on a 1.2% agarose gel.

To amplify a subsequence of cytochrome oxidase subunit I we redesigned the primers C1-J-1751 (alias Ron) and C1-N-2191 (alias Nancy) of Simon *et al*. (1994) based on two sequences, *Mus musculus domesticus* (Murinae, Accession no AB042432, Mori *et al*. unpublished) and *Sorex unguiculatus* (Soricidae, Accession no. AB061527, Nikaido *et al*. 2001). These two species were the only representatives within the families of Murinae, Soricidae or Arvicolinae, for which mitochondrial cytochrome oxidase I sequences were available in GenBank. The new, partly degenerate, primers were

as follows: RonM (5′-GGMGCMCCMGATATRGCATTCCC-3′), NancyM (5′-CCTGGGAGRATAAGAATATAWACTTC-3′). Amplification of cytochrome oxidase I was prepared using the same composition of DNA, primer and HotStar Taq™ MasterMix (Qiagen) as for cytochrome *b*, however, PCR was performed on different thermal cyclers (Techne, models Genius or Progene), generally running 45 cycles of 40 s at 95 °C, 40 s at 50 °C and 1 min at 72 °C. The initial denaturation step at 95 °C lasted 15 min, the final extension at 72 °C lasted 7 min.

Double-stranded PCR products were purified with the QIAquick PCR purification kit following the protocol of the manufacturer (Qiagen). The amount of DNA was measured with a BioPhotometer (Eppendorf). Linear amplification was performed using the purified PCR product directly. Each sequencing reaction (8 μL) consisted of 30 ng double-stranded PCR product, 2 μL BigDye™ Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and 4 μM of each primer. Sequencing was performed on an ABI Prism 3100 Genetic Analyser (Applied Biosystems). Complementary sequences from each individual were visually checked and all genotypes aligned with the help of the program SEQUENCHER v4.0.5 (Gene Codes Corp.).

Sequence analysis

To identify our specimen on the basis of cytochrome *b*, an extensive selection of published sequences from mouse, vole and shrew species existing in Switzerland was added to our sequences to be run in a cluster analysis (Table 2). Cytochrome *b* sequences were analysed from the first coding base to base 795 of the gene compared with the corresponding coding gene sequences of *Mus musculus domesticus* (Accession no. AY057807, Lundrigan *et al*. 2002). To evaluate the usefulness of cytochrome oxidase I sequences, as an alternative or additional identification tool for small

Table 2 To confirm field identification of our specimen, cytochrome *b* (cyt *b*) sequences of different species within the taxa of mice, voles and shrews existing in Switzerland were extracted from GenBank to be included in a cluster analysis. A bootstrap consensus tree of a NJ cluster analysis is presented in Fig. 1. *Murinae were excluded from this analysis for concise presentation

GenBank Accession no. cyt *b*	Species	Alias	Country
Insectivora, Soricidae			
AJ312033 (Brünner *et al*. 2002)	*Sorex araneus*	Sorara	Switzerland
AJ312032 (Brünner *et al*. 2002)	*Sorex araneus*	Sorara	France
AJ312030 (Brünner *et al*. 2002)	*Sorex araneus*	Sorara	United Kingdom
AJ312028 (Brünner *et al*. 2002)	*Sorex araneus*	Sorara	Hungary
AJ312035 (Brünner *et al*. 2002)	*Sorex antinorii*	Sorant	Italy
AJ312036 (Brünner *et al*. 2002)	*Sorex antinorii*	Sorant	Switzerland
AJ000419 (Brünner *et al*. 2002)	*Sorex coronatus*	Sorcor	France
AJ000431 (Fumagalli *et al*. 1999)	*Sorex minutus*	Sormin	Switzerland
AJ000447 (Fumagalli *et al*. 1999)	*Sorex alpinus*	Soralp	Switzerland
AJ000446 (Fumagalli *et al*. 1999)	*Sorex alpinus*	Soralp	Switzerland
AJ000465 (Fumagalli *et al*. 1999)	*Neomys fodiens*	Neofod	Serbia
AJ000466 (Fumagalli *et al*. 1999)	*Neomys anomalus*	Neoano	Italy
AJ000467 (Fumagalli *et al*. 1999)	*Crocidura russula*	Crorus	Switzerland
AB077280 (Iwasa & Ohdachi unpubl.)	*Crocidura suaveolens*	Crosua	Austria
Rodentia, Arvicolinae		Clegla	Germany
AF159401 (Martin *et al*. 2000)	*Clethrionomys glareolus*	Clegla	Germany
AF159400 (Martin *et al*. 2000)	*Arvicola terrestris*	Arvter	Switzerland
AF159403 (Martin *et al*. 2000)	*Microtus arvalis*	Micarv	Germany
AY167212 (Jaarola & Searle 2002)	*Microtus agrestis*	Micagr	Czech Rep.
AY167211 (Jaarola & Searle 2002)	*Microtus agrestis*	Micagr	Sweden
AF159402 (Martin *et al*. 2000)	*Microtus agrestis*	Micagr	Switzerland
Rodentia, Murinae*		Aposyl	Germany
AF159395 (Martin *et al*. 2000)	*Apodemus sylvaticus*	Aposyl	Germany
AJ311148 (Michaux *et al*. 2002)	*Apodemus sylvaticus*	Aposyl	Italy
AJ311149 (Michaux *et al*. 2002)	*Apodemus sylvaticus*	Aposyl	France
AJ311151 (Michaux *et al*. 2002)	*Apodemus flavicollis*	Apofla	France
AJ311150 (Michaux *et al*. 2002)	*Apodemus flavicollis*	Apofla	Italy
AF159392 (Martin *et al*. 2000)	*Apodemus flavicollis*	Apofla	Germany
AJ311153 (Michaux *et al*. 2002)	*Apodemus alpicola*	Apoalp	Austria
AF159391 (Martin *et al*. 2000)	*Apodemus alpicola*	Apoalp	Switzerland
AY057807 (Lundrigan *et al*. 2002)	*Mus musculus domesticus*	Musmus	Switzerland

mammals, we used the coding bases from sites 285–723 (site positions being compared to *Mus musculus domesticus*; Accession no. AB042432, Mori *et al*. unpublished). Sequences were aligned in SEQUENCHER v4.0.5.

Unrooted neighbour-joining (NJ) trees were constructed for cytochrome *b* and cytochrome oxidase I with MEGA v2.1 (Kumar *et al*. 2001) using the Kimura 2-parameter distance (Kimura 1980), including transversions and transitions unweighted and handling gaps by pairwise deletion. As we were only interested in species clustering and not in deeper phylogenetic signals, we included all bases although we are aware that third codon position transitions may be saturated (Martin *et al*. 2000). Confidence in the estimated relationship was determined by bootstrap analyses using both distance and parsimony methods. Figure 1 shows the two NJ bootstrap consensus trees from 1000 replicates, bootstrap values being posted above branches. In addition, bootstrap values based on unweighted maximum parsimony (MP) are posted at the base of species clades below branches. MP was estimated using PAUP 4.0b10 (Swofford 2003), doing a heuristic search using simple stepwise addition and performing tree-bisection–reconnection (TBR) branch swapping.

Microarray-based diagnostics

Oligonucleotide probes 21–27 bp in length were designed manually on an alignment including all cytochrome oxidase I sequences presented in Table 1. All selected probes were tested for thermodynamic properties using Vector NTI 8 (InforMax, Inc.), and the two most appropriate probes were chosen for each of the following seven species:

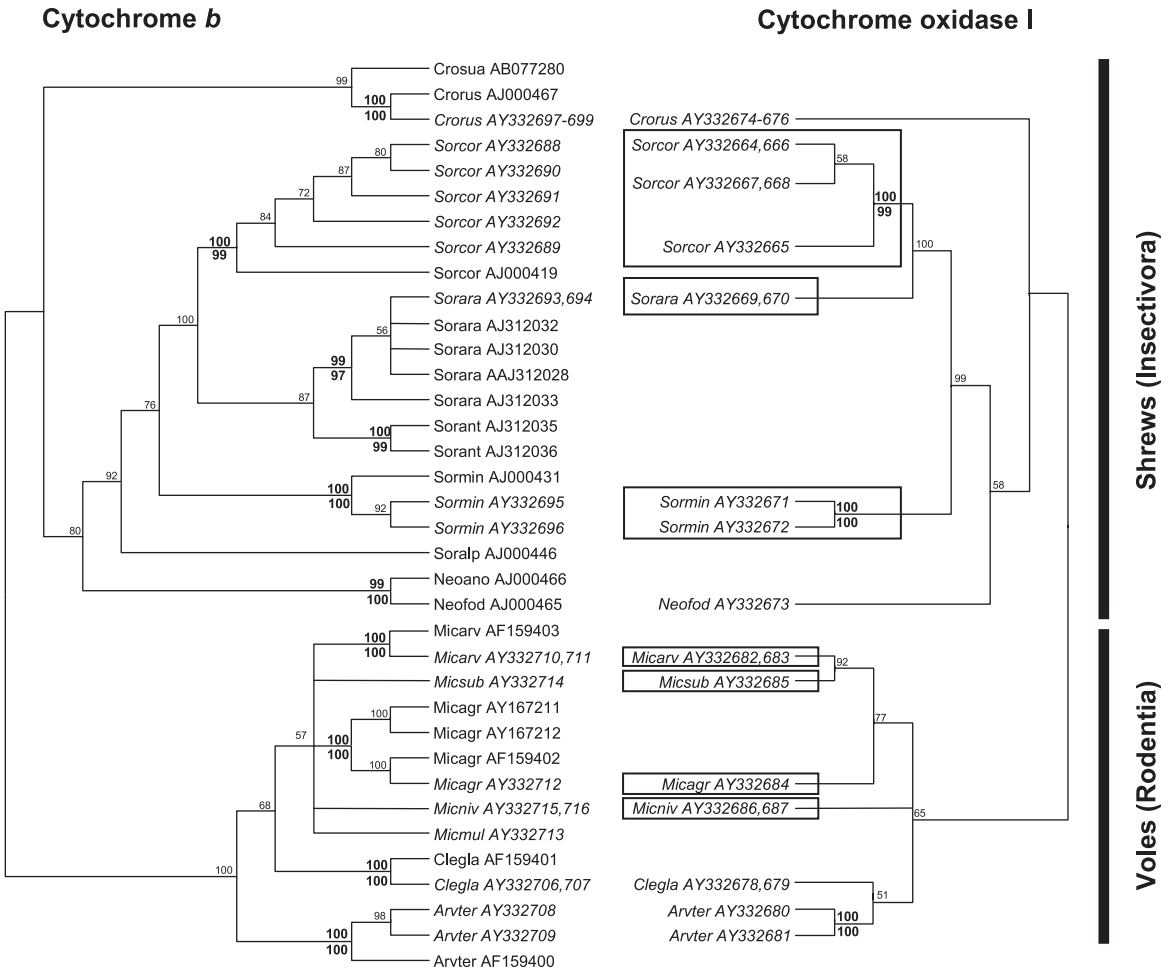

Fig. 1 Bootstrap consensus trees of cytochrome *b* and cytochrome oxidase I displaying the unambiguous discrimination of different species of voles and shrews. The trees are constructed using the neighbour-joining (NJ) method performed on a Kimura 2-parameter corrected distance matrix. Numbers above branches represent bootstrap values from NJ (1000 replications), bootstrap values below branches have been calculated by maximum parsimony reconstruction (10 000 replications). Bootstrap values for species clades are written in bold. Haplotypes from this study are named in italics, haplotypes from GenBank in normal letters. Framed haplotypes were used for the implementation of a microarray-chip.

Table 3 Oligonucleotide probes designed for microarray based diagnostics. For each species, two different probes from the cytochrome oxidase I fragment were chosen. Tm = melting temperature

Probe-name	Species	Sequence (5′– to 3′, 3′-aminolinker)	GC (%)	Tm (°C)	Length (bp)
Insectivora, Soricidae					
Sorcor1	*Sorex coronatus*	TAGGGGTGTTTGATATTGAGATATAGC	37.0	53.5	27
Sorcor2	*Sorex coronatus*	AATGCCAGCTGCAAGAACTGG	52.4	56.7	21
Sorara1	*Sorex araneus*	GGATGAAACGCCTGCTAAATGAAG	45.8	58.0	24
Sorara5	*Sorex araneus*	GTGTTTGATATTGAGACATAGCAGGG	42.3	55.1	26
Sormin2	*Sorex minutus*	AATTGATTGAGCCGAGAATTGATG	37.5	56.2	24
Sormin5	*Sorex minutus*	TAATTCCGGCTGCAAGGACAG	52.4	56.6	21
Rodentia, Arvicolinae					
Micarv3	*Microtus arvalis*	CTGGTGGTTTCATATTGATGATAGTGG	40.7	56.8	27
Micarv4	*Microtus arvalis*	TGTTAAGAGTATTGTGATCCCTGCG	44.0	56.4	25
Micagr2	*Microtus agrestis*	GAATTGATGATACACCCGCTAAGTGT	42.3	56.6	26
Micagr4	*Microtus agrestis*	TTGGTATTGTGTCATTGCTGGTG	43.5	54.9	23
Micsub2	*Microtus subterraneus*	CTATTGATGATGCCAGAAGTAAGAGGA	40.7	56.2	27
Micsub5	*Microtus subterraneus*	GGAGAGAAGAAGGAGTACCGCAGT	54.2	56.2	24
Micniv2	*Microtus nivalis*	GGAGGGAGAAAATAGTTAGATCGACTG	44.4	56.3	27
Micniv3	*Microtus nivalis*	TACTGGTAGCGAGAGAAGTAGGAGG	52.0	54.5	25

S. araneus, *S. coronatus*, *S. minutus*, *Microtus arvalis*, *M. agrestis*, *M. subterraneus* and *M. nivalis*. The specifications of the probes were rather tolerant with 37–54.2% GC content and a melting temperature (Tm) between 53.5 and 58 °C (Nearest Neighbour, constant parameters of DNA and salt concentration, Table 3).

Fifty microlitres of 3′-6C-aminoterminated probes (Microsynth) were diluted 1:1 in Genetix Microarray Spotting Solution to a final concentration of 25 μM and spotted onto aldehyde-coated microarray slides (Genetix) using a QArrayMini automated spotter (Genetix). Each slide included three replicates of each block, one block including two adjacent replicates of each probe and two different probes for each species, as well as a standard, a cy3-labelled amino-linked probe (Fig. 2). Each spot was stamped four times and spotting was conducted in 75% relative humidity. After spotting, the slides were washed following the protocol for aldehyde-coated slides of the manufacturer.

PCR products for hybridization were labelled using the same protocol and primers as for the amplification of cytochrome oxidase I, but in addition, the forward primer RonM was marked with a cy3-fluorophore at the 5′-end. Amplified double-stranded PCR products were denaturated for 10 min at 96 °C, cooled on ice for 3–5 min and then kept on ice until hybridized. For hybridization, the 439 bp PCR products were used directly without fragmentation or purification. Hybridization was conducted in a Lucidea Slidepro (Amersham Biosciences) injecting a mixture of total volume of 200 μL, containing 15 μL of the unpurified PCR product, 85 μL H_2O and 100 μL genHYB Microarray Hybridization Buffer (Genetix). After priming the slide with 2 × SSC, the DNA was hybridized at 40 °C for 2 h, followed by three washes at low stringency at 32 °C (6 min in 2 × SSC and 0.2% SDS, 2 min in 0.2 × SSC and 0.2% SDS and 2 min in 0.075 × SSC). To test the reproducibility and specificity of the probes, PCR products from all individuals with sequenced cytochome oxidase I (Table 1) were hybridized at low stringency onto the chip, including additional five species which were not represented on the chip but which were included in probe design. The slides were analysed in a GenePix Personal 4100 A microarray scanner (Axon Instruments) at a wavelength of 532 nm with maximal laser power to detect even weak cross hybridizations.

Results

Sequence analysis

Cytochrome *b* subsequences were successfully sequenced from 29 individuals of 14 different species (Table 1, AY332688–AY332716) resulting in 23 unique haplotypes. Sequencing of the cytochrome oxidase I gene was successful for 24 individuals from 12 different species (Table 1, AY332664–AY332687) and resulted in 16 haplotypes. Difficulties occurred while trying to sequence cytochrome *b* as well as cytochrome oxidase I from *Apodemus* sp. No sequence of a functional cytochrome oxidase I fragment could be obtained from *A. sylvaticus* or *A. flavicollis*. Instead, a single gene copy that would not translate into a functional protein was found and suspected to be a pseudogene. Cytochrome *b* sequences could be produced for both *Apodemus* species, however, there was always some polymorphism observed. One to two polymorphic sites were also encountered in the sequences of *Microtus arvalis* and *M. agrestis*. *Neomys fodiens* was only amplified under difficulties and only for cytochrome oxidase I. For phylogenetic analyses sequences

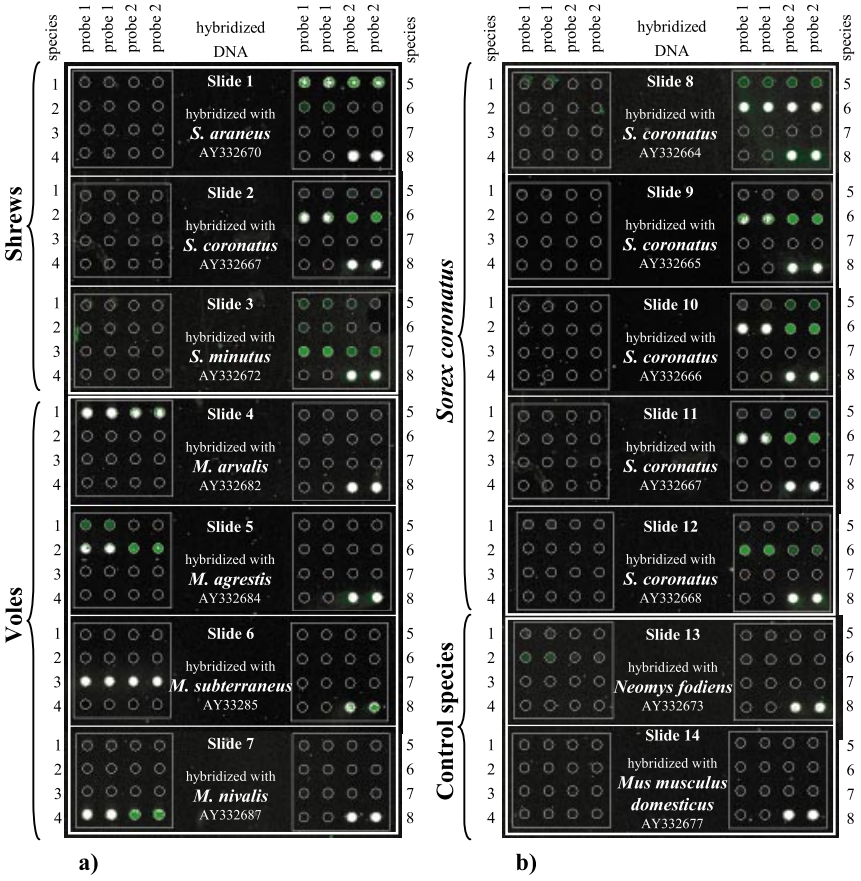

Fig. 2 Analysis of voles and shrews using microarray technology. Each slide illustrates the hybridization with cytochrome oxidase I amplicons from one individual. The probes were arranged in three identical blocks on each slide, but only one block is shown here. Each block includes all probes for all species. Each species is represented by a row, each row consisting of two probes that were spotted in an adjacent replicate to a total of four spots. *Microtus* probes are placed in rows 1–4 in a square of 4 × 4 spots on the left (1, *M. arvalis*; 2, *M. agrestis*; 3, *M. subterraneus*; 4, *M. nivalis*), *Sorex* probes in rows 5–7 on the right (5, *S. araneus*; 6, *S. coronatus*; 7, *S. minutus*). Row 8 represents a negative control (H_2O) and a standard (cy3-labelled probe), each again in replicate. (a) Slides 1–7: one representative individual per species (b) Slides 8–12: five individuals of *S. coronatus* to illustrate the variation within species; Slides 13 and 14: two exemplary test species that are not represented by probes on the chip.

from Murinae (*Apodemus* sp. and *Mus musculus domesticus*) were excluded.

Phylogenetic analysis (NJ and MP) of cytochrome *b* included an alignment of 36 haplotypes, 17 haplotypes from this study (Table 1, Fig. 1) and 19 sequences from GenBank (Table 2). The alignment produced 799 characters with 344 variable sites. The average nucleotide composition was 28.6% T, 28.2% C, 29.3% A and 13.9% G.

The alignment of the cytochrome oxidase I sequences for phylogenetic analysis was based on 15 haplotypes (Table 1, Fig. 1), with 439 characters, 172 of them being variable. The sequences were composed of an average of 31.0% T, 28.3% C, 25.8% A and 14.9% G.

Figure 1 shows the resulting bootstrap consensus trees from NJ analysis. Both, the cytochrome *b* and cytochrome oxidase I tree form distinct species clades that are supported in at least 99% of all bootstrapped trees (Fig. 1, bootstrap values above branches, in bold). A second analysis based on MP resulted in comparable bootstrap values (Fig. 1, bootstrap values below branches).

Microarray-based diagnostics

Cytochrome oxidase I was used as a marker for the development of a microarray chip. Two probes per species were designed, and all gave a specific signal for the seven species tested (Fig. 2). Absolute signal intensities varied among individuals as well as among different probes within individuals, and the second probe generally showed lower absolute intensities than the first. Cross-hybridization (false positives) could be observed in several species (e.g. *Sorex araneus*, *S. minutus*, *Microtus agrestis*), however, the signal intensities of the true positives were always distinctly stronger. The maximal ratio of false to true signals for any probe within a species (mean signal intensity over all replicates and all individuals) was < 4% for *M. arvalis*, *M. subterraneus* and *M. nivalis*, 12 ± 7% (± SD) for *S. araneus*, 12 ± 14% for *S. coronatus*, 32 ± 12% for *M. agrestis* and 57 ± 23% for *S. minutus*. The maximal observed ratio within a single individual was 77 ± 15% (DNA from *S. minutus*, AY332671, probe 'sormin5'-compared with probe 'micarv3'; mean over all replicates).

The sequence differences between a specific probe and the DNA sequences of the other species included in the alignment was between 2 and 8 bases. The number of base differences had a significant effect on the mean ratio of false to correct positives (ANOVA: $P > 0.002$): The mean ratio of false to true signals (± SD) for probes with a 2 bp difference was 13 ± 10%, for probes with 3 bp difference 6.5 ± 4.7% and for probes with 4–9 bp differences below 6% in

descending order. One exception with a mean ratio of 6.5% were the data from probes with 8 bp difference, which included the above-mentioned cross-hybridization of the DNA from one *S. minutus* individual with *M. arvalis* probes.

The five species that were not represented on the chip but were included in the alignment for probe design (*Neomys fodiens*, *Crocidura russula*, *Mus musculus domesticus*, *Clethrionomys glareolus*, *Arvicola terrestris*) showed nought to only low levels of cross-hybridization (two examples in Fig. 2).

Discussion

Comparison of the 439 bp sequence of cytochrome oxidase subunit I with the 795 bp sequence of cytochrome *b* showed that both gene fragments discriminate well among species of voles and shrews in cluster analysis, for example, between the closely related species *Microtus arvalis* and *M. agrestis* or *Sorex araneus* and *S. coronatus*. None of the haplotypes tested formed a clade with any species other than their own. It is important to note that it is discrimination among species that is important for diagnostics, whereas deeper phylogenetic signals are irrelevant for our purposes.

Hair sampling is frequently used in studies of larger carnivores, particularly for DNA-based population inventories of larger mammals such as bears or for species that are hard to see or trap such as pine martens (Paetkau 2003). However, to our knowledge no studies have included non-invasive collection methods for molecular diagnostics in biodiversity monitoring studies of small mammals. In general, DNA and proteins are extracted from tissue (liver, heart, kidney, muscle, ear, toe). The advantages of hair sampling are obvious: animals do not have to be harmed and threatened species are not at risk. Another advantage of this approach is the possibility of using hairs from hair traps, especially when dealing with small carnivores such as weasels, stouts or martens, or arboreal species such as dormice. In this study, extraction of DNA from hair samples generally produced satisfactory results. However, with *Apodemus* sp., as well as *Microtus* sp., we were confronted with polymorphic copies originating from more than one sequence when amplifying the two mitochondrial genes. This is a possible consequence of using degenerate primers that may also amplify nuclear pseudogenes (DeWoody *et al*. 1999). This problem of using degenerate primers has to be addressed carefully in developing microarray diagnostics based on mtDNA. For cases in which nuclear pseudogenes must be assumed, as in *Apodemus*, optimized primers need to be designed and it has to be verified that the resulting PCR fragment is of mitochondrial origin, for example, by comparing the sequence with the amplification product obtained from RNA extraction.

Our approach for a diagnostic microarray-chip is straightforward and simple. It represents a low-cost method in terms of time and money, important factors for the application of a new method particularly in the field of ecology and agriculture. Once a chip has been designed and is available, any small laboratory should be able to carry out the diagnostics in a short time. All that is needed is equipment for PCR and a scanner (or a fluorescence microscope) to analyse the fluorescently labelled slides. Hybridization can be done using cover slips and hybridization chambers rather than a hybridization instrument. Compared with other diagnostic methods such as sequencing or DNA-restriction fragment length polymorphism (RFLP), microarrays offer a reduction in the number of work steps as well as time (no linear amplification, no further purification steps, no gels, etc.). In addition, no replicates are generally needed as each array includes several redundant features as well as replicates among and within blocks. However, microarrays are not necessarily the best tool for large-scale screening. As each analysis needs one slide, alternative methods such as PCR-RFLP based on highly parallel work steps (e.g. including robotics) might, in this case, be considered and rated more favourable in terms of time and money.

The design of this 'proof-of-principle' chip was done by visually choosing probes, which were surprisingly accurate and robust. This might be the easiest approach for a low-density chip including up to 10–20 species. For chips including more species or more haplotypes, however, bioinformatics will be unavoidable for probe design. In addition, probe design should include a three- to fivefold redundancy including probes from different gene regions per species to adequately discriminate between unknown species. To achieve this robustness it may be necessary to extend probe design to several genes, and cytochrome oxidase I is an eligible candidate for a complement to cytochrome *b*.

Direct hybridization of double-stranded amplicons gave sufficient discrimination even among closely related species such as *S. araneus* and *S. coronatus*, although we observed relatively high variation among individuals and among probes in terms of their hybridization efficiency, which may be partly due to the nonstandardized amount of target DNA applied. Interestingly, the second probe showed a lower absolute intensity than the first in all seven species tested. The second probes were generally designed more towards the 3'-end of the target DNA than were the first probes. Hybridization of single-stranded amplicons did not improve the intensity of the second probes. Therefore, we conclude that the observed differential efficiency is not an effect of incomplete denaturation, but might be the result of physical position of the hybridized strand towards the slide surface.

Our study showed that cytochrome oxidase I has the same discrimination power as cytochrome *b* for species identification in mice, voles and shrews. Shrews belong to the order of Insectivora, which comprises 440 species world wide (9% of all mammalian species, Nowak 1999),

mice and voles to the order of Rodentia with over 2000 species (43% of all mammalian species, Nowak 1999). Successful species identification of sibling species within these groups makes the cytochrome oxidase I gene a promising alternative marker for the species identification of mammals. As our chip illustrates, cytochrome oxidase I as a complement to the mammalian cytochrome *b* gene is useful, whether used in a mixed-phyla microarray such as the hypothetical 'Biodiversity-Chip', or as an additional marker to enhance redundancy, for example within a 'Mammalia-Chip'.

Acknowledgements

We thank Michel Blant, Tiziano Maddalena and Paul Marchesi for their collaboration in the BDM Z3 monitoring including trapping and sample donations, Urs Käser and Andrea Pfunder for the 'what the cat brings home'-sampling, Beatrice Frey and Sandra Jakob for their assistance in the laboratory, and Loren Rieseberg and three anonymous reviewers for their valuable suggestions on an earlier version of the manuscript. This work was partly supported by funding from COST Action 853 and the BDM Z3 programme of SAEFL.

References

Bodrossy L, Stralis-Pavese N, Murrell JC *et al.* (2003) Development and validation of a diagnostic microbial microarray for methanotrophs. *Environmental Microbiology*, **5**, 566–582.

Boonham N, Walsh K, Smith P *et al.* (2003) Detection of potato viruses using microarray technology: towards a generic method for plant viral disease diagnosis. *Journal of Virological Methods*, **108**, 181–187.

Brünner H, Lugon-Moulin N, Balloux F, Fumagalli L, Hausser J (2002) A taxonomical re-evaluation of the Valais chromosome race of the common shrew *Sorex araneus* (Insectivora: Soricidae). *Acta Theriologica*, **47**, 245–275.

Bystřická D, Lenz O, Mráz I, Dědič P, Šíp M (2003) DNA microarray: parallel detection of potato viruses. *Acta Virologica*, **47**, 41–44.

Churchfield S (1990) *The Natural History of Shrews*. Christopher Helm, London.

DeWoody JA, Chesser RK, Baker RJ (1999) A translocated mitochondrial cytochrome *b* pseudogene in voles (Rodentia: *Microtus*). *Journal of Molecular Evolution*, **48**, 380–382.

Fumagalli L, Taberlet P, Stewart DT *et al.* (1999) Molecular phylogeny and evolution of *Sorex* shrews (Soricidae: Insectivora) inferred from mitochondrial DNA sequence data. *Molecular Phylogenetics and Evolution*, **11**, 222–235.

Gurnell J, Flowerdew JR (1994) *Live Trapping Small Mammals: A Practical Guide*, 2nd edn. The Mammal Society, London.

Hausser J, ed. (1995) *Säugetiere der Schweiz*. Denkschriftenkommission der Schweizerischen Akademie der Naturwissenschaften. Birkhäuser-Verlag, Basel.

Haynes S, Jaarola M, Searle JB (2003) Phylogeography of the common vole (*Microtus arvalis*) with particular emphasis on the colonization of the Orkney archipelago. *Molecular Ecology*, **12**, 951–956.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. *Proceedings of the Royal Society London Series B*, **270**, 313–321.

Hintermann U, Weber D, Zangger A, Schmill J (2002) Biodiversitäts-Monitoring Schweiz BDM. *Zwischenbericht, Hrsg, Bundesamt für Umwelt, Wald und Landschaft BUWAL. Schriftenreihe Umwelt* no. 342.

Irwin DM, Kocher TD, Wilson AC (1991) Evolution of the cytochrome *b* gene of mammals. *Journal of Molecular Evolution*, **32**, 128–144.

Jaarola M, Searle JB (2002) Phylogeography of field voles (*Microtus agrestis*) in Eurasia inferred from mitochondrial DNA sequences. *Molecular Ecology*, **11**, 2613–2621.

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, **16**, 111–120.

Kingsley MT, Straub TM, Call DR *et al.* (2002) Fingerprinting closely related *Xanthomonas* pathovars with random nonamer oligonucleotide microarrays. *Applied and Environmental Microbiology*, **68**, 6361–6370.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis. Vers. 2.1. *Bioinformatics*, **17**, 1244–1245.

Lundrigan BL, Jansa SA, Tucker PK (2002) Phylogenetic relationships in the genus *Mus*, based on paternally, maternally, and biparentally inherited characters. *Systematic Biology*, **51**, 410–431.

Martin Y, Gerlach G, Schlötterer Ch, Meyer A (2000) Molecular phylogeny of European muroid rodents based on complete cytochrome *b* sequences. *Molecular Phylogenetics and Evolution*, **16**, 37–47.

Mazurok NA, Rubtsova NV, Isaenko AA *et al.* (2001) Comparative chromosome and mitochondrial DNA analyses and phylogenetic relationships within common voles (*Microtus*, Arvicolidae). *Chromosome Research*, **9**, 107–120.

Michaux JR, Chevret P, Filippucci M-G, Macholan M (2002) Phylogeny of the genus *Apodemus* with a special emphasis on the subgenus *Sylvaemus* using the nuclear IRBP gene and two mitochondrial markers: cytochrome *b* and 12*S* rRNA. *Molecular Phylogenetics and Evolution*, **23**, 123–136.

Michaux JR, Kinet S, Filippucci M-G *et al.* (2001) Molecular identification of three sympatric species of wood mice (*Apodemus sylvaticus*, *A. flavicollis*, *A. alpicola*) in western Europe (Muridae: Rodentia). *Molecular Ecology Notes*, **1**, 260–263.

Mitchell-Jones AJ, Amori G, Bogdanowicz W *et al.* (1999) *The Atlas of European Mammals*. Poyser, London.

Nikaido M, Kawai K, Cao Y *et al.* (2001) Maximum likelihood analyses of the complete mitochondrial genomes of eutherians and a reevaluation of the phylogeny of bats and insectivores. *Journal of Molecular Evolution*, **53**, 508–516.

Nowak RM (1999) *Walkers Mammals of the World*, 6th edn. Johns Hopkins University Press, Baltimore.

Paetkau D (2003) An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology*, **12**, 1375–1387.

Riddle AE, Pilgrim KL, Mills LS, McKelvey KS, Ruggiero LF (2003) Identification of mustelids using mitochondrial DNA and non-invasive sampling. *Conservation Genetics*, **4**, 241–243.

Rubenstein K (2003) Commercial aspects of microarray technology. *Biotechniques*, **34**, 52–54.

Rudi K, Katla T, Naterstad K (2003) Multi locus fingerprinting of *Listeria monocytogenes* by sequence-specific labeling of DNA probes combined with array hybridization. *FEMS Microbiology Letters*, **220**, 9–14.

Simon C, Frati F, Beckenbach AT *et al.* (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Stewart DT, Perry ND, Fumagalli L (2002) The maritime shrew, *Sorex maritimensis* (Insectivora: Soricidae): a newly recognized Canadian endemic. *Canadian Journal of Zoology*, **80**, 94–99.

Stine OC, Carnahan A, Singh R *et al.* (2003) Characterization of microbial communities from coastal waters using microarrays. *Environmental Monitoring and Assessment*, **81**, 327–336.

Swofford DL (2003) PAUP*. *Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MAs.

Taberlet P, Meyer A, Bouvet J (1992) Unusual mitochondrial DNA polymophism in two local populations of blue tit (*Parus caerulens*). *Molecular Ecology*, **1**, 27–36.

Vogel P, Maddalena T, Mabille A, Paquet G (1991) Confirmation biochimique du statut spécifique du mulot alpestre *Apodemus alpicola* Heinrich, 1952 (Mammalia, Rodentia). *Bulletin de la Société Vaudoise Des Sciences Naturelles*, **80**, 471–481.

Wang D, Coscoy L, Zylberberg M *et al.* (2002) Microarray-based detection and genotyping of viral pathogens. *Proceedings of the National Academy of Sciences of the USA*, **99**, 15687–15692.

Wilson DE, Cole FR, Nichols JD, Rudran R, Foster MS (1996) *Measuring and Monitoring Biology Diversity: Standard Methods for Mammals*. Smithsonian Institution Press, Washington, DC.

Zardoya R, Meyer A (1996) Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution*, **13**, 933–942.

Juerg E. Frey is head of the laboratory for molecular diagnostics at Agroscope FAW Wädenswil (Swiss Federal Research Station for Horticulture) and chair of the COST action 853 (Agricultural Biomarkers for Array Technology). Monika Pfunder and Jürg E. Frey are working on microarray chips to identify economically relevant agricultural pests. Otto Holzgang works at the Swiss Ornithological Institute, his projects among others include wildlife and traffic as well as monitoring programmes (i.e. brown hare, grey partridge). He and M. Pfunder are also involved in the biodiversity monitoring module Z3 of the Swiss Agency for the Environment, Forests and Landscape (SAEFL), covering mammals.
