
Biological Identifications Through DNA Barcodes

Hassan A. I. Ramadan and Nabih A. Baeshen

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/49967>

1. Introduction

Although much biological research depends upon species diagnoses, taxonomic expertise is collapsing. We are convinced that the sole prospect for a sustainable identification capability lies in the construction of systems that employ DNA sequences as taxon 'barcodes'. It was established previously that the mitochondrial gene cytochrome *c* oxidase I (COI) can serve as the core of a global bio-identification system for animals. A new tools were developed recently to be complementary markers for (COI) DNA barcoding.

Species identification is essential in food quality control procedures or for the detection and identification of animal material in food samples. Recent food scares e.g. avian flu and swine flu, malpractices of some food producers and religious reasons have tremendously reinforced public awareness regarding the composition of food products. However, because labels do not provide sufficient guarantee about the true contents of a product, it is necessary to identify and/or authenticate the components of processed food, thus protecting both consumers and producers from illegal substitutions [1]. In addition, trade of endangered species has contributed to severe depletion of biodiversity.

Numerous analytical methods that rely on protein analysis have been developed for species identification, such as electrophoresis techniques [2], immunoassays [3] and liquid chromatography [4]. However, these methods are of limited use in species identification. The progress of molecular biology introduced a new approach, which is based on nucleotide sequence diversities among species in particular regions of DNA [5–7]. The nucleotide regions chosen for species identification were varied by researchers. Within vertebrates, a cytochrome *b* (cyt *b*) gene in the mitochondrial DNA has been studied from multiple viewpoints including the nucleotide diversity among species [6] and the availability of nucleotide sequence data for references [5]. Many of the other regions studied are also located in the mtDNA. The coding regions for 12S and 16S ribosomal RNA [8–10], and the noncoding D-loop region [7, 11, 12] have shown their potential to be the targets for the species test.

Although central to much biological research, the identification of species is often difficult. DNA sequencing, with key sequences serving as a pattern “barcode”, has therefore been proposed as a technology that might expedite species identification [13].

DNA barcoding promises fast, accurate species identifications by focusing analysis on a short standardized segment of the genome [14]. Several studies have now established that sequence diversity in a 650-bp fragment of the mitochondrial gene cytochrome c oxidase I (cox1; also referred to as COI) provides strong species-level resolution for varied animal groups including birds [15], fishes [16] and Lepidoptera [17].

Besides the cox1 gene, other mitochondrial markers also have been widely sequenced across vertebrates for their utility in phylogenetic or to complement cox1 in DNA barcoding.

In amphibians the 16S ribosomal RNA gene (16S) has been suggested as a complementary DNA barcoding marker [18]. Another protein coding gene, cytochrome b, has also been suggested as a marker to determine species boundaries [19, 20].

An attempt was made to present a phylogenetic systematic framework for an improved barcoder as well as a taxonomic framework for interweaving classical taxonomy with the goals of ‘DNA barcoding’ [21]. Another study showed that DNA arrays and DNA barcodes are valuable molecular methods for biodiversity monitoring programs [22]. In this chapter we introduce the use of specific fragments of mitochondrial ribosomal RNA from Egyptian buffalo to be used as a perfect barcode for identification of closely related species. Also, we will extend this study to include distantly species identification [23-24]. Our studies were also extended for chickens and small organisms like mites to be studied by both nuclear and mitochondrial markers. Identification of these mites is very important for biological control programs.

All these methods could be used for global bio-identification system or forensic science development.

2. Materials and methods

2.1. DNA purification

Genomic DNA was extracted from peripheral blood of Egyptian buffalo's and chickens by using standard commercial Kit (Pure-gene Genomic DNA purification Kit) as recommended by the manufacturer (www.gentra.com). In case of mites, Genomic DNA was extracted using Capture Column kit method, total DNA was purified using generation DNA purification system.

2.2. Primers used for amplifications of PCR specific fragments

2.2.1. D-loop primers

These primers yielded a PCR product of 1142 base pairs. This encompasses the whole of the D-loop and includes flanking sequence at both ends [12].

IL0500: 5'AGGCATTTTCAGTGCCTTGC-3'
 IL0501: 5'TAGTGCTAATAACCAACGGCC-3'

Two additional new forward primers (SH-1 and SH-2) specific for buffalo were designed inside the D-loop sequence to facilitate sequencing and correction processes.

SH-1: 5' CCT CGC ATG TAC GGC ATA CA-3'
 SH-2: 5'CAA CCC TTC AGG CAA GGA TC-3'

2.2.2. Primers used for amplification of specific fragments from mites

Two target DNA fragments of the predatory mite, *A. swirskii* were PCR amplified and sequenced: a fragment in the central part of the mitochondrial cytochrome oxidase subunit I gene (COI) and the fragment of the nuclear ribosomal transcribed spacers (ITS) [25-26]. The COI primers were designed specifically for tetranychid mites. They were:

5'TGATTTTTTGGTCACCCAGAAG3' and 5'TACAGCTCCTATAGATAAAAAC 3'.

The ITS region was amplified using the primers 5'AGAGGAAGTAAAAGTCGTAACAAG 3' for the 3' end of 18S rDNA and 5' ATATGCTTAAATTCAGGGG 3' for the 5' end of the 28S.

2.2.3. Primers used for amplification of the first 539 base fragment of the D-loop region of the birds

The conserved primer pair, L16750 (forward; 5'-AGG ACT ACG GCT TGA AAA GC-3') and H 547 (reverse; 5'- ATG TGC CTG ACC GAG GAA CAA G-3') were used to amplify the first 539 base fragment of the D-loop region of the birds. The primer number refers to the positions of the 3' end of the primer in the reference sequence [27].

2.2.4. 12S primers

Primers specific for mitochondrial 12S rRNA gene were synthesized [23]:

5'-CAAAGTGGGATTAGATACCCCACTAT-3'; 5'-AGGGTGACGGGCGGTGTGT-3' and directed towards the two conserved regions of the gene. The primers were synthesized by Amersham Pharmacia Biotech (U.K.).

2.2.5. 16S primers

PCR amplification and direct sequencing With two universal primers (sense, 5'-GTGCAAAGGTAGCATAATCA-3' and antisense, 5'-TGTCCTGATCCAACATCGAG-3') directed toward conserved regions [24], the polymerase chain reaction was used to amplify homologous segments of mitochondrial 16S rRNA from four animal species belonging to family Bovidae, including river buffalo, cattle, sheep and goat.

2.3. The amplification reaction

The amplification reaction used for amplification of the D-loop fragment was also used (with little modifications in temperature cycling) in the other experiments according to the conditions of each experiment.

The amplification reaction was carried out in a 25 μ l reaction mixture consisting of 1.25 unit Taq polymerase (DyNAzyme), 1X enzyme buffer (1X is 10 mM Tris-HCl, pH 8.8 at 25 °C, 1.5 mM MgCl₂, 50 mM KCl and 0.1% Triton X-100) supplied by the manufacture, 1 μ M of each forward and reverse primer, 0.2 mM dNTPs and 100 ng of DNA. The reaction mixture was overlaid with sterile mineral oil and was run in an MJ research PTC-100 Thermocycler. The temperature cycling was as follows: 30 cycles of 45 seconds at 94°C; 1 minute at 58°C and 1 minute at 71°C, followed by a final extension at 71°C for 5 minutes. All PCR amplifications included a negative control reaction which lacked template DNA. No product was seen in any negative control. Small quantities of the reaction products (5 μ l each) were used for electrophoresis with an appropriate size marker on 1.5% agarose in 1X-Tris acetate buffer (TAE).

After electrophoresis the gels were stained with ethidium bromide and were examined with UV lamp at a wave length 312 nm to verify amplification of the chosen specific fragment. The PCR products were purified using QIAquick PCR purification kit (Qiagen, Inc.) and the resulting purified products were used in the subsequent sequencing reactions. Sequencing was performed on an Applied Biosystems 310 genetic analyzer (Applied Biosystem) using Big Dye terminator cycle sequencing ready reaction mixture according to manufacturer's instructions (Applied Biosystems).

2.4. Sequence analysis and multiple sequence alignment

Pairwise sequence alignments were carried out using NCBI-BLASTN 2.2.5 version & PSI BLAST. Multiple sequence alignments were done using the MUSCLE 3.6 software and CLUSTALW (1.82). Analysis, manipulation, conservation plots, positional entropy plot and conserved region analysis was done using the BIOEDIT package. Variable sites were extracted from the multiple sequence alignment using the MEGA 3.1 package [12].

2.5. Phylogenetic analysis

Phylogenetic model selection was done using the FINDMODEL server available from the HCV LANL database at (<http://hcv.lanl.gov/content/hcv-db/findmodel/>). A Bayesian phylogenetic tree was constructed by Markov chain Monte Carlo (MCMC) method as implemented in the MR BAYES 3.1 package using the Hasegawa-Kishino-Yano plus Gamma model HKY+G substitution model with an invariant four category gamma distribution among sites. A 50% consensus tree was generated and the analysis was repeated two times. Maximum parsimony tree was conducted using MEGA version 4, with 1000 bootstraps for reliability.

The mean overall, within group and between groups genetic distances were done using the MEGA 4.0 software [12].

3. Results

Our experience in the field of molecular identification or DNA barcoding through a series of published research papers are represented in this section. Results with some illustrated figures and tables are represented here but the complete information could be obtained through obtaining the complete published papers from the publication section.

Shows the Positional entropy plot of the D-loop for the buffalo, and cow sequences. The Bayesian phylogenetic trees of cow and buffalo sequences were constructed using MRBAYES software (Figure 2) and Maximum parsimony tree using the Kimura two-parameter model and the closest neighbor interchange method of the MEGA 3.1 software package (Figure 3). Table 1. Shows the Substitution events detected in complete D-loop sequences from multiple sequence alignments between cows and buffaloes.

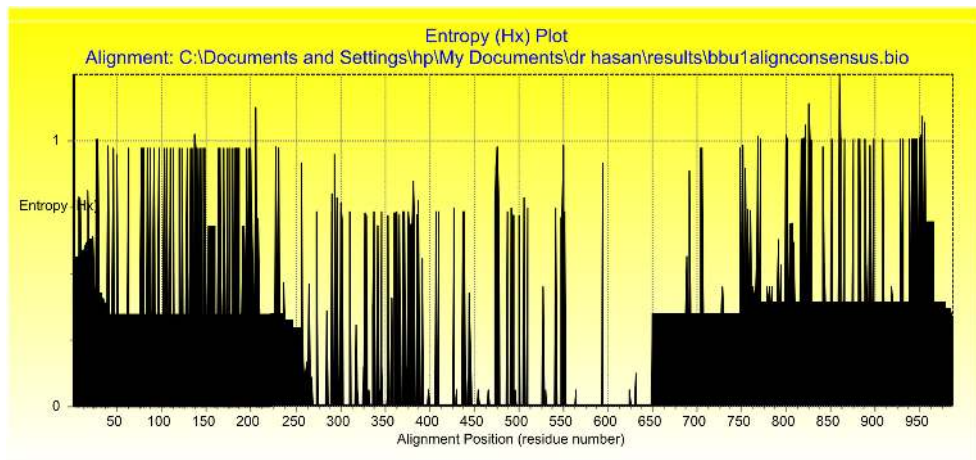


Figure 1. Positional Entropy Plot of the D loop for the Buffalo, and cow sequences. The X axis is the nucleotide position, while the Y axis is the entropy (lack of information) at that position. The central region is mainly conserved, and the beginning and end regions are highly variable (shaded areas).

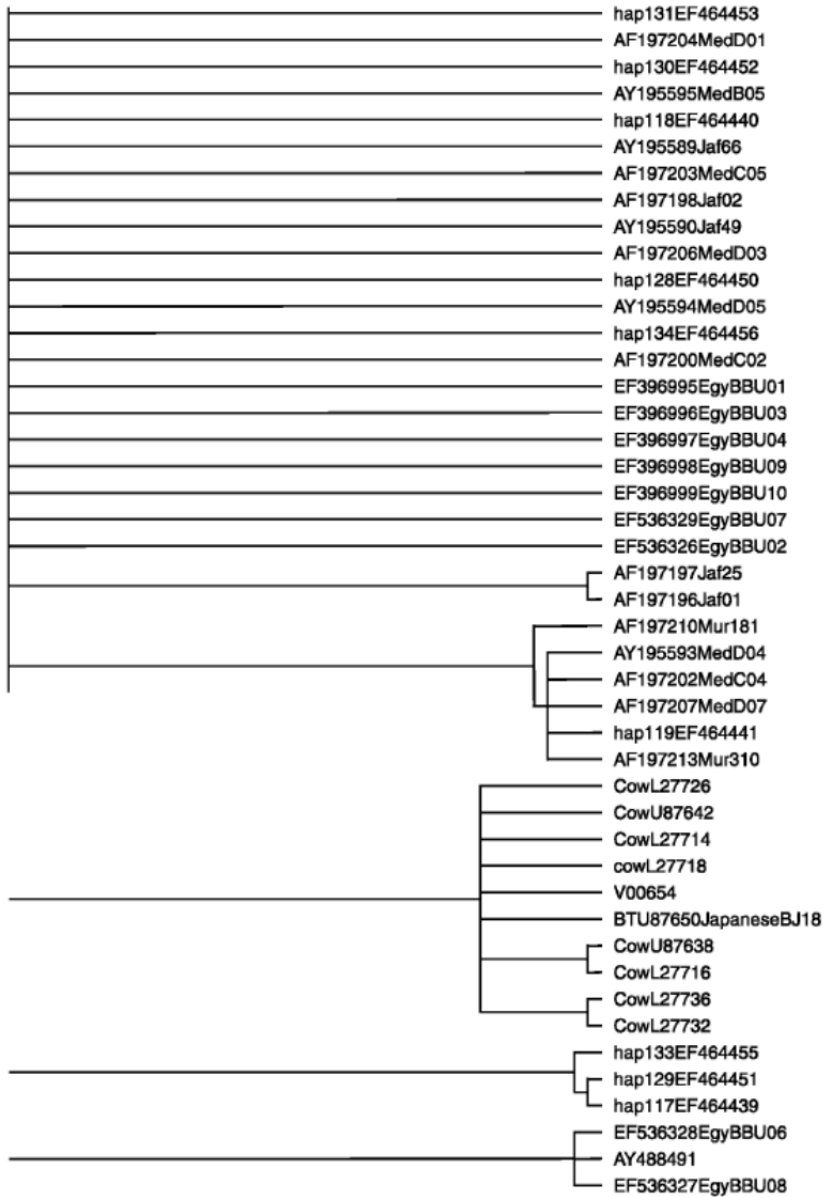


Figure 2. Bayesian phylogenetic tree of the cow and buffalo sequences using the MRBAYES software.

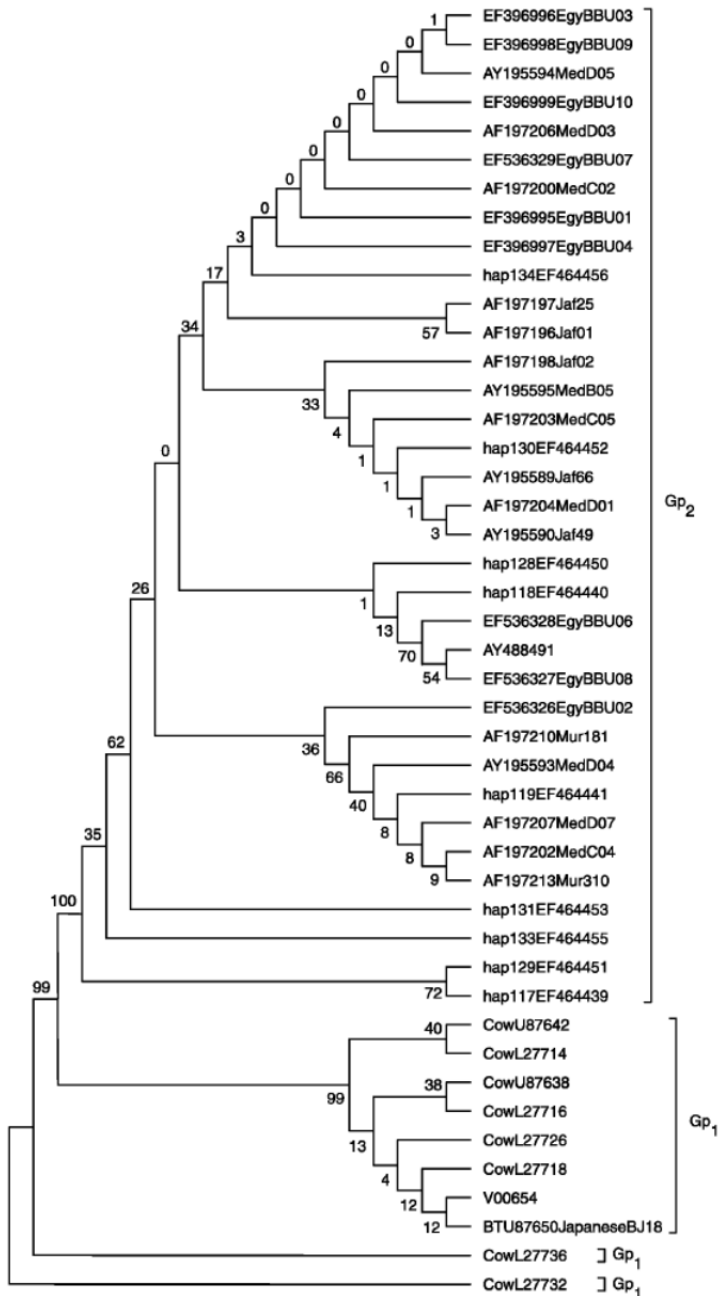


Figure 3. Maximum parsimony tree constructed using the Kimura two-parameter model and the closest neighbor interchange method of the MEGA 3.1 software package. The numbers show the percentage of bootstrap confidence.

Animals	Cows (35)		Buffaloes (53)		Cows & buffaloes (88)	
	Value	S E	Value	S E	Value	S E
Substitutions and genetic distances						
Total numbers of transitions	44		20		100	
Total numbers of transversions	4		1		61	
Total number of indels (insertions/deletions)	1		7		15	
Total number of substitutions	49		28		176	
R ratio (transversions/transitions)	0.09		0.05		0.61	
Genetic distance within group	0.023	0.003	0.007	0.002		
Genetic distance between the two groups					0.156	0.016
Overall (all animals) distance					0.06	0.006

Table 1. Substitution events detected in complete D-loop sequences from multiple sequence alignment between Cows (35 animals) and buffaloes (53 animals).

Shows the PCR amplification of chicken mitochondrial D loop fragments while the phylogenetic tree constructed between the Egyptian and GenBank database chicken samples is represented in Figure 5. The Polymorphic sites and their positions are shown in Table 2.

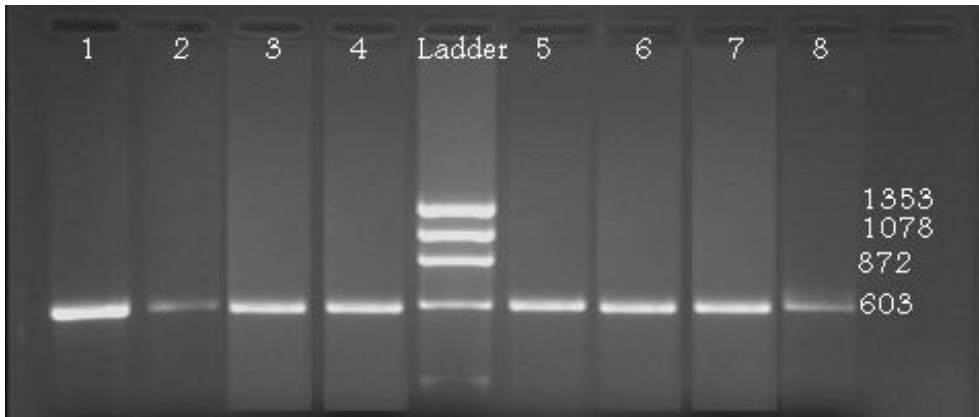


Figure 4. PCR amplification of chicken mitochondrial D loop fragment. The PCR reactions were run on 1% agarose gel, stained with ethidium bromide and examined with UV. Samples from 1 to 4 for Dandarawi breed and samples from 5 to 8 for Fayoumi breed. The *Hae* III digest of Φ X174 DNA was used as ladder (1353, 1078, 872, and 603 base pairs).

output2.ph Sat Dec 29 00:51:38 2007 Page 1 of 1



Figure 5. Phylogenetic tree constructed between the Egyptian and GenBank database chicken samples. Sample DQ629875 was used as out-group for its high diversity.

Breed & Accession number	Variable sites and their positions					
	23	35	276	457	464	483
EgyDand1 EF586881	T	A	C	A	G	T
EgyDand2 EF586882	T	A	C	C	T	T
EgyDand3 EU352856	T	A	C	A	G	A
EgyFay1 EF586879	T	A	A	A	G	A
EgyFay2 EF586880	T	A	C	C	T	A
DQ629875	A	*	C	A	G	A
Database public sequence	T	A	C	A	G	A

Table 2. The Polymorphic sites and their positions. The nucleotide positions were given with respect to the Egyptian nucleotide numbers in GenBank database. The left column shows the breeds with their accession numbers.

The following Figures: show polyacrylamide gel representing the PCR-amplified fragment of CO1 and its sequence (Figures 6 & 8) while the PCR-amplified fragment of ITS region and its sequence were presented in Figures 7 & 9.

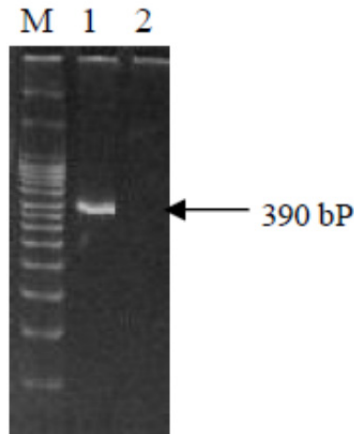


Figure 6. 10% polyacrylamide gel representing the PCR-amplified product of CO1. M: 50 bp DNA size marker, lane 1: CO1 fragment and lane 2: blank (PCR cocktail without DNA).

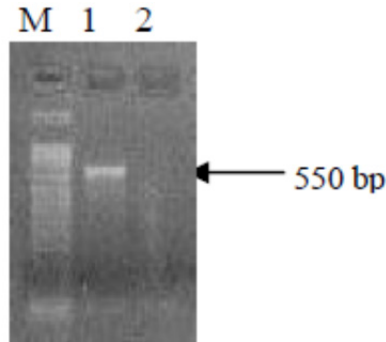


Figure 7. Agarose gel representing the PCR-amplified product of ITS. M: 50 bp DNA size marker, lane 1: ITS fragment and lane 2: blank.

```
5'CAATCTAATAAGTTTCAGACTTTCGGGGCTTTGGGTATAATTTATGCTATATTGTCTATTGGG
ATTTTAGGGTTTATTGTATGAGCTCATCATATTTACAGTAGGAATAGATGTTGATTCTCGAG
CTTATTTTACCGCTGCTACGATAATTATTGCGGTACCTACAGGTATTAAGGTTTTTCTTGATT
ATCTACAATGTATGGGTCAGTAGTAAAGTGAGGAGTAGTTGTTTTGTGAAAATTTGGTTTTATT
TTTTTATTCAGGTTAGGTGGTTTAACCGGGTGGTTTTGTCAAA3'
```

Figure 8. Sequence of the mitochondrial gene fragment coding for cytochrome oxidase subunit I(COI) in the predacious mite, *Amblyseius swirskii*.

```
5'GNCACCTGATTTAGATCACGTCATAATGGTTTGTGTTTGGTTACTCTTGTGACCAAGAGAATT
TATACCATTGCATGGCGATACCGACGAGGGCTGCAGCAAAGGTAATACGTATGATACGGTT
TATATACTCGAAACAAATGTAGCACAGATACAGTAAGTGTCCGTGCTGAAATTTTCATTCAA
AAACACAATGCTCACAAAATTTACAACCTCACATCAATTTCCACAAATTAATGTTTTTCAT
CGATTTAAGGACTAAGTGATCCCCCATATTGAGTCTTGGTTTTTTCTTCTCGATAGCACAACT
TACTTCCAAAGGAAAGTGAAAAGGTTTGTCCGAATGGTTACCGACTTACTGTGCGATACGCCTT
CCCCGTAACCAAAAAGGGACCGTTAAACACCCACCAATTCGAAGCGGTTGTCGCCCAAAA
GGGACCGGTTTGGAAAGACCGTTATCCACCGTTAAAACCTAAATTACAGGTTTGTTTTTTCAT
TGGTCTTTTTTCGAAAAAACACAN3
```

Figure 9. Sequence of the internal transcribed spacers (ITS) region in the predacious mite, *Amblyseius swirskii*. N: unknown base.

Sequencing of a fragment of ITS region (ITS1, ITS2, 5.8S) indicated almost complete identity of the Egyptian samples with *Neoseiulus swirskii*, accession number EU 310505 (= *A. swirskii*). Regardless the locality within Egypt, taxa were less identical when compared to the related species *N. Cucumeris* (Oud.) *N. andersoni* (Chant), and *N. fallacis* (Gar-man).

According to the molecular analysis (Table 3), samples are grouped into three groups.

Nucleotide number (Accession No. EU924213)			2	3	3	8	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	0	3	3	8	3	3
			3	3	5	8	5	6	7	8	9	3	2	6	3	9	2	1	4	7	0	4	3	4	5	3	3	2
Sample	Accession Number	Phylogenetic Group	ITS1	5.8S rRNA													ITS2											
Sample 3	EU924213	Group 1	c	c	c	a	c	c	a	t	a	c	c	a	g	t	c	a	t	a	c	c	-	a	g	-	a	g
Sample 4	EU924214		c	c	c	a	c	c	a	t	a	c	c	a	g	t	c	a	t	a	c	c	-	a	g	-	a	g
Sample 2	EU924215		a	a	t	a	t	c	a	t	a	-	c	a	g	t	c	a	t	a	c	c	-	a	g	-	a	g
Sample 1	EU924212	Group 3	a	a	t	-	t	g	t	-	-	-	t	-	t	t	t	t	-	t	-	t	t	t	t	t	a	-
<i>N. swirskii</i>	EU310505		a	a	t	-	t	g	t	-	-	-	t	-	t	t	t	t	-	t	-	t	t	t	t	t	a	t
Sample 5	EU924216	Group 2	n	a	t	-	t	g	t	-	-	-	t	a	g	a	t	t	-	c	c	t	t	a	t	g	g	t
Sample 6	EU924217		n	n	n	-	t	g	t	-	-	-	t	a	g	a	t	a	-	c	-	t	t	a	t	-	g	t

Table 3. The variable sites (a = Adenine, c = Cytosine, g = Guanine, t = Thymine, - = deletion and n = Not detected) detected in a fragment of nuclear ITS region of six samples of *A. swirskii* collected from citrus and grapes in the Nile delta of Egypt.

The results of 12S rRNA showed that, two haplotypes of 12S rRNA sequences were identified from the multiple alignment results between the nine tested Egyptian buffalo sequences and other examples of homologous buffalo sequences selected from GenBank database. Two buffalo haplotypes were revealed, of which haplotype 1 which include Egyptian buffaloes and haplotype 2 which include Chinese swamp buffalo; breed: Haikou (accession AY702618), Mediterranean (accession AY488491) and *Bubalus bubalis* (accession AF231028). The detected SNPs can be classified as shown in table 4.

Eleven SNPs were detected which can be used to discriminate between subfamily Bovinae, represented by buffalo and cattle and the subfamily Caprinae represented by sheep and goat.

No. of SNPs	Specificity	Representing nucleotide	Base position
8	all buffaloes (haplotypes 1 & 2)	Guanine	at positions 110, 132 and 196
		Thymine	position 172
		Cytosine	positions 71, 269, 271 and 348.
3	buffalo haplotype 1 only	Guanine	position 158
		Cytosine	position 267
		Adenine	position 293
4	buffalo haplotype 2 only	Thymine	positions 32 and 267
		Guanine	position 293
		Cytosine	position 72
12	cattle (<i>Bos taurus</i>)	Guanine	positions 193, 266 and 273
		Adenine	27 and 174
		Thymine	positions 26, 36, 186 & 190
		Cytosine	positions 253, 294 and 295.
6 + 2 indels	sheep (<i>Ovis aries</i>)	Guanine	positions 172, 231 and 300
		Adenine	position 190
		Thymine	position 271
		Cytosine	position 349
		insertion of Adenine	position 164-165
		deletion of Adenine	position 259
6 + 1 deletion	goat (<i>Capra hircus</i>)	Guanine	position 299
		Thymine	positions 260, 261 and 317
		Cytosine	positions 225 and 259
		deletion of Adenine	deletion at position 273

Table 4. Nucleotide variation in a specific 12S rRNA gene fragment of four studied Bovidae species. Nucleotide positions correspond to Egyptian buffaloes GenBank accession numbers (FJ828575-FJ828583).

Considering multiple alignment results between homologous 16S rRNA sequences obtained from GenBank database with the reference sequence, it was shown that, the entire 16S rRNA fragment (422 bp. in size) contains more than 57 variable sites (from base no. 21 to base no. 323) inside the two conserved regions. The bases outside this variable region are completely conserved in the four species (Figure 10 and Table 5). From these variable sites, 25 specific nucleotides were chosen (which gave clear significant results in both types of alignment comparisons (two and multiple alignment sequences programs) as a reference for identification of unknown species (from base no. 21 to base no. 308). It was also shown that the size of the amplified fragments were less by one nucleotide (421 bp) in case of goat and two nucleotides (420 bp) in case of both cattle and sheep.

Detection of specific variable sites between Egyptian buffalo 16S rRNA gene fragment and the other studied three species is shown to be a good marker for identification of the four studied species. The detected variable sites can be classified as represented in both Fig. 10 and Table 5.

EGY BBU reference sequences FJ748599 - FJ748607	Variable site	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
	Base No.	21	22	33	55	55	11	11	11	11	11	11	11	11	11	11	22	22	22	22	22	22	22	22	23	23	33
	16S rRNA	C	T	G	T	A	T	A	G	T	C	A	C	G	C	C	C	T	A	A	A	T	G	C	T	C	
AF492350 Bosi	Cattle	T	C	A	T	T	A	G	T	C	A	T	A	A	C	C	C	T	A	A	A	T	A	C	T	C	
AB074968 JBC8																											
AB074967 JBC7																											
AB074966 JBC6																											
AB074965 JBC5																											
AB074964 JBC3																											
AB074963 JBC2																											
AB074962 JBC1																											
AY526085 Bost																											
AF492351 Bost																											
AY126697 Bosi																											
V00654 Bost																											
DQ867009 BBU		River Buffalo					A										C	C									
FJ748599 Egy BBU																											
FJ748600 Egy BBU																											
FJ748601 Egy BBU																											
FJ748602 Egy BBU																											
FJ748603 Egy BBU																											
FJ748604 Egy BBU			C	T	G											C	G										
FJ748605 Egy BBU																											
FJ748606 Egy BBU																											
FJ748607 Egy BBU																											
AY488491 BBU																											
AY327817 BBU																											
AF547270 BBU																											
DQ867013 Ovis	Sheep	T	C	A			C		G	A	C		T	A				C					A	A		T	
AY858379 Ovis									G	A	C		T	G					G				C		C		
AF010406 Ovis																											
DQ334809 Capra	Goat							A	G			C	A					T							T		
AF533441 Capra																			A								
M55541 Capra																											
DQ867011 Capra																											

Figure 10. Multiple sequence alignments result showing the total variable sites between river buffalo (BBU), Cattle (Bost, Bosi and JBC), Sheep (Ovis) and goat (Capra) in the specific 16S rRNA fragment. Sequences of the 16S rRNA fragment of Egyptian buffaloes (FJ748599–FJ748607). Differing nucleotides are noted (T, A, G and C)

No. of Variable sites (SNPs)	Specificity	Representing nucleotides	Base position
Six	River buffaloes	Guanine	(36, 189, and 297)
		Thymine	(29)
		Cytosine	(21 and 182)
Three	Cattle (<i>Bos taurus</i>)	Guanine	(55)
		Thymine	(190 and 221)
Seven	Sheep (<i>Ovis aries</i>)	Guanine	(119, 171, and 251)
		Adenine	(122)
		Thymine	(167)
		Cytosine	(295 and 301)
Four	Goat (<i>Capra hircus</i>)	Guanine	(252 and 254)
		Thymine	(227)
		Cytosine	(50).
Five	Group one (river buffaloes and cattle, Subfamily Bovinae)	Thymine	(102, 129 and 249)
		Cytosine	(298 and 308)
Five	Group two (sheep and goat, Subfamily Caprinae).	Adenine	(298)
		Cytosine	(102, 129 and 249)
		Thymine	(308)

Table 5. Nucleotide variation in a specific 16S rRNA gene fragment of four studied Bovidae species. Nucleotide Positions correspond to Egyptian buffaloes GenBank accession numbers (FJ748599–FJ748607)

4. Discussion

4.1. DNA barcoding, genome evolution & phylogenetic trees

The ability of molecular trees to encompass both short and long periods of time is based on the observation that different genes evolve at different rates. The DNA specifying ribosomal RNA (rRNA) changes relatively slowly, so comparisons of DNA sequences in these genes are useful for investigating relationships between taxa that diverged hundreds of millions of years ago. Studies of the genes for rRNA have shown, for example, that fungi are more closely related to animals than to green plants—something that certainly could not have been deduced from morphological comparisons alone.

In contrast, the DNA in mitochondria (mtDNA) evolves relatively rapidly and can be used to investigate more recent evolutionary events.

The methodology used in DNA barcoding has been straightforward. Sequences of the barcoding region are obtained from various individuals. The resulting sequence data are then used to construct a phylogenetic tree using a distance-based 'neighbour-joining' method. In such a tree, similar, putatively related individuals are clustered together. The term 'DNA barcode' seems to imply that each species is characterized by a unique sequence,

but there is of course considerable genetic variation within each species as well as between species. However, genetic distances between species are usually greater than those within species, so the phylogenetic tree is characterized by clusters of closely related individuals, and each cluster is assumed to represent a separate species.

An evolutionary tree (or *phylogenetic tree*) is a branching diagram that represents the evolutionary history of a group of organisms. For example, we might use morphological and genetic data to figure out a phylogenetic tree of animals. Such a tree can provide a huge amount of information. For any particular group of animals our tree could identify the ancestors and closest relatives of the group. If we traced the history of animals all the way back, we could use the tree to help us answer questions such as, What did the earliest animals look like? What features did they pass on to all their descendants?. Phylogenetic trees also have great practical value. The same techniques we use to reconstruct evolutionary history have been used in forensics, where phylogenetic trees have helped solve criminal cases, and epidemiology, where trees have been used to estimate when and where diseases such as AIDS originated.

Now that we can compare entire genomes, including our own, some interesting facts have emerged. As you may have heard, the genomes of humans and chimpanzees are strikingly similar. An even more remarkable fact is that homologous genes are widespread and can extend over huge evolutionary distances. While the genes of humans and mice are certainly not identical, 99% of them are detectably homologous. And 50% of human genes are homologous with those of yeast.

It is not a coincidence that DNA barcoding has developed in concert with genomics-based investigations.

DNA barcoding (a tool for rapid species identification based on DNA sequences) and genomics (which compares entire genome structure and expression) share an emphasis on large scale genetic data acquisition that offers new answers to questions previously beyond the reach of traditional disciplines. DNA barcodes consist of a standardized short sequence of DNA (400–800 bp) that in principle should be easily generated and characterized for all species on the planet (1). A massive on-line digital library of barcodes will serve as a standard to which the DNA barcode sequence of an unidentified sample from the forest, garden, or market can be matched. Similar to genomics, which has accelerated the process of recognizing novel genes and comparing gene function, DNA barcoding will allow users to efficiently recognize known species and speed the discovery of species yet to be found in nature. DNA barcoding aims to use the information of one or a few gene regions to identify all species of life, whereas genomics, the inverse of barcoding, describes in one (e.g., humans) or a few selected species the function and interactions across all genes.

To be practical as a DNA barcode a gene region must satisfy three criteria: (i) contain significant species-level genetic variability and divergence, (ii) possess conserved flanking sites for developing universal PCR primers for wide taxonomic application, and (iii) have a short sequence length so as to facilitate current capabilities of DNA extraction and

amplification. A short DNA sequence of 600 bp in the mitochondrial gene for cytochrome *c* oxidase subunit 1 (CO1) has been accepted as a practical, standardized species-level barcode for animals (see www.barcoding.si.edu). The inability of CO1 to work as a barcode in plants set off a race among botanists to find a more appropriate marker. A number of candidate gene regions have been suggested as possible barcodes for plants, but none have been widely accepted by the taxonomic community. This lack of consensus is in part due to the limitations inherent in a plastid marker relative to plant CO1, and also because a quantitative context for selecting a gene region as a barcode for plants has not been offered. Several factors must be considered and weighted in selecting a plant DNA barcode: (i) universal PCR amplification, (ii) range of taxonomic diversity, (iii) power of species differentiation, and (iv) bioinformatics analysis and application.

4.2. Molecular genetics reveals evolutionary relationships

Evolution results from the accumulation of inherited changes in populations. Because DNA is the molecule of heredity, evolutionary changes must be reflected in changes in DNA. Systematics have long known that comparing DNA within a group of species would be a powerful method for inferring evolutionary relationships, but for most of the history of systematics, direct access to genetic information was nothing more than a dream. Today, however, **DNA sequencing**—determining the sequence of nucleotides in segment of DNA – is comparatively cheap, easy, and widely available. *The polymerase chain reaction* (PCR) allows systematics to easily accumulate large samples of DNA from organisms, and automated machinery makes sequence determination a comparatively simple task.

4.3. Direct benefits of DNA barcoding undoubtedly include

- i. make the outputs of systematics available to the largest possible community of end-users by providing standardized and high-tech identification tools, e.g. for biomedicine (parasites and vectors), agriculture (pests), environmental assays and customs (trade in endangered species);
- ii. relieve the enormous burden of identifications from taxonomists, so they can focus on more pertinent duties such as delimiting taxa, resolving their relationships and discovering and describing new species;
- iii. pair up various life stages of the same species (e.g. seedlings, larvae);
- iv. provide a bio-literacy tool for the general public.

Perhaps another advantage of DNA barcoding is that it will also facilitate basic biodiversity inventories. Indeed, from the premises of molecular phylogenetics to assembling the tree of life, DNA sequences in environmental sampling and reconstruction of phylogenetic trees to place sequences into an evolutionary context have been used in several inventories of cryptic biodiversity (e.g. soil bacteria or marine/freshwater micro-organisms).

New 'Genetic Bar Code' Technique Establishes Ability to Derive DNA Information from RNA

Science Daily (Apr. 8, 2012) — Researchers from Mount Sinai School of Medicine have developed a method to derive enough DNA information from non-DNA sources -- such as RNA -- to clearly identify individuals whose biological data are stored in massive research repositories. The approach may raise questions regarding the ability to protect individual identity when high-dimensional data are collected for research purposes.

A paper introducing the technique appears in the April 8 online edition of *Nature Genetics*.

DNA contains the genetic instructions used in the development and functioning of every living cell. RNA acts as a messenger that relays genetic information in the cell so that the great majority of processes needed for tissue to function properly can be carried out.

To date, access to databases with DNA information has been restricted and protected as it has long been considered the sole genetic fingerprint for every individual. However, vast amounts of RNA data have been made publicly available via a number of databases in the United States and Europe. These databases contain thousands of genomic studies from around the world.

In this study, authors developed a technique whereby a person's DNA could be inferred from RNA data using gene-expression levels monitored in any of a number of tissues. In contrast, most studies involving DNA and RNA begin with DNA sequences and then seek to associate expression patterns with changes in DNA between individuals in a population. This is the first time going from RNA levels to DNA sequence has been described.

"By observing RNA levels in a given tissue, we can infer a genotypic barcode that uniquely tags an individual in ways that enables matching the individual to an independently derived DNA sample.". Not only can genotypic barcodes be deduced from RNA, but RNA levels in some tissue can inform not only individual characteristics like age and sex, but on diseases such as Alzheimer's and cancer, as well as the risks of developing those diseases."

Author details

Hassan A. I. Ramadan* and Nabih A. Baeshen
*Department of Biological Sciences, Faculty of Science,
King Abdulaziz University, Jeddah, Saudi Arabia*

Hassan A. I. Ramadan
Department of Cell Biology, National Research Centre, Dokki, Cairo, Egypt

Acknowledgement

The authors gratefully acknowledge the financial support from the Deanship of Scientific Research (DSR) at King Abdulaziz University (KAU) represented by the Unit of Strategic Technologies Research through the Project number **(10-BIO1257-03)**.

* Corresponding Author

5. References

- [1] Pascal G, Mahe S (2001) Identity, traceability, acceptability and substantial equivalence of food. *Cell Mol Biol* 47:1329–1342
- [2] Skarpeid HJ, Kvaal K, Hildrum KI (1998) Identification of animal species in ground meat mixtures by multivariate analysis of isoelectric focusing protein profiles. *Electrophoresis* 19:3103–3109
- [3] Hsieh YH, Sheu SC, Bridgman RC (1998) Development of a monoclonal antibody specific to cooked mammalian meats. *J Food Prot* 61(4):476–487
- [4] Ashmoor SH, Monte WC, Stiles PG (1998) Liquid chromatographic identification of meats. *J Assoc Off Anal Chem* 71:397–403
- [5] Parson W, Pegoraro K, Niederstatter H, Fögger M, Steinlechner M (2000) Species identification by means of the cytochrome b gene. *Int J Legal Med* 114:23–28
- [6] Hsieh HM, Chiang HL, Tsai LC, Lai SY, Huang NE, Linacre A, Lee JC (2001) Cytochrome b gene for species identification of the conservation animals. *Forensic Sci Int* 122:7–18
- [7] Murray BW, McClymont RA, Strobeck C (1995) Forensic identification of ungulate species using restriction digests of PCR amplified mitochondrial DNA. *J Forensic Sci* 40(6):943–951
- [8] Balitzki-Korte B, Anslinger K, Bartsch C, Rolf B (2005) Species identification by means of pyrosequencing the mitochondrial 12S rRNA gene. *Int J Legal Med* 119:291–294
- [9] Rodríguez MA, García T, González I, Asensio L, Hernández PE, Martín R (2004) PCR identification of beef, sheep, goat, and pork in raw and heat-treated meat mixtures. *J Food Prot* 67(1):172–177
- [10] Rodríguez MA, García T, González I et al (2003) Identification of goose, mule, duck, chicken, turkey, and swine in foie gras by species-specific polymerase chain reaction. *J Agric Food Chem* 51:1524–1529
- [11] Montiel-Sosa JF, Ruiz-Pesini E, Montoya J, Roncales P, Lopez-Pérez MJ, Pérez-Martos A (2000) Direct and highly species-specific detection of pork meat and fat in meat products by PCR amplification of mitochondrial DNA. *J Agric Food Chem* 48:2829–2832
- [12] Ramadan HAI, El Hefnawi M (2008) Phylogenetic analysis and comparison between cow and buffalo (including Egyptian buffaloes) mitochondrial displacement-loop regions. *Mitochondrial DNA* 19(4):401–410
- [13] Waugh J (2007) DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays* 29(2):188–197
- [14] Hebert PDN, Cywinska A, Ball SL, de Waard JR (2003) Biological identifications through DNA barcodes. *Proc R Soc B* 270:313–321
- [15] Hebert PDN, Stoeckle MY, Zemplak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biol* 2:E312
- [16] Ward RD, Zemplak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philos Trans R Soc Lond B* 360:1847–1857

- [17] Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proc Natl Acad Sci USA* 103:968–971
- [18] Vences M, Thomas M, Meijden AVD, Chiari Y, Vieites DR (2005) Comparative performance of the 16S rRNA gene in DNA barcoding of amphibians. *Front Zool* 2:5. doi:10.1186/1742-9994-2-5
- [19] Bradley RD, Baker RJ (2001) A test of the genetic species concept: cytochrome b sequences and mammals. *J Mammal* 82: 960–973
- [20] Lemer S, Aurelle D, Vigliola L, Durand JD, Borsa P (2007) Cytochrome b barcoding, molecular systematics and geographic differentiation in rabbitfishes (Siganidae). *C R Biol* 330:86–94
- [21] De Salle R, Egan MG, Siddall M (2005) The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos Trans R Soc Lond B* 360(1462):1905–1916
- [22] Hajibabaei M, Singer GAC, Clare EL, Hebert PDN (2007) Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. *BMC Biol* 5:24. doi:10.1186/1741-7007-5-24)
- [23] Ramadan HAI, Mahfouz ER (2009) Sequence of specific mitochondrial 12S rRNA fragment of Egyptian buffalo as a reference for discrimination between buffalo, cattle, sheep and goat. *J Appl Biosci* 21:1258–1264
- [24] Ramadan, H.A.I (2011). Sequence of specific mitochondrial 16S rRNA gene fragment from Egyptian buffalo is used as a pattern for discrimination between river buffaloes, cattle, sheep and goats. *Mol. Biol. Rep.* 38 (6) 3929-3934.
- [25] Ramadan, H.A.I., El-Banhawy, E.M., Hassan, A.A., and Afia,S.I. (2004): Genetic variation in the predacious phytoseiid mite, *Amblyseius Swirskii* (Acari: Phytoseiidae): Analysis of specific mitochondrial and nuclear sequences. *Arab J. Biotech.*, 7, No.(2): 189-196.
- [26] Ramadan, H. A. I, El-Banhawy, E. M and Afia, S. I. (2009). On the identification of taxa collected from Egypt in the species sub-group *andersoni* Chant and McMurtry: Morphological relationships with related species and molecular analysis of inter and intra-specific variations (Acari: Phytoseiidae). *Acarologia* (France), Vol 49, fasc 3-4: 115-120
- [27] Ramadan, H.A.I, Galal, A., Fathi, M.M., El Fiky, S.A., Yakoub, H.A. (2011): Characterization of Two Egyptian Native Chicken Breeds Using Genetic and Immunological Parameters. *Biotechnology in Animal Husbandry* 27 (1), p 1-16.