The Journal of Basic and Applied Zoology

RESEARCH Open Access

Cytochrome b shows signs of adaptive protein evolution in *Gerbillus* species from Egypt

Mahmoud Amin Khalifa[1*], Mahmoud Ibrahim Younes[2] and Ahmed Ghazy[1]

Abstract

Background: Amino acid polymorphisms in the mitochondrial cytochrome *b* (cyt*b*) gene of four *Gerbillus* species have been investigated for their geographical distribution and possible functional significance. The sequences were obtained from a total of 20 specimens representing four species of genus *Gerbillus* collected from Siwa Oasis, Dabaa, Wadi El Natron, El Faiyum, and Baltim in Egypt.

Results: Our results identified a group of amino acid variant polymorphisms that were useful for both species taxonomic and biogeographic assignments. The results demonstrated that amino acid variants L>F173 (Leucine>Phenylalanine), A>M203 (Alanine>Methionine), and I>V221(Isoleucine>Valine) were specific to *G. andersoni*, while the variant V>M283 (Valine>Methionine) was only specific to *G. andersoni* from Baltim. The variants, L>P263 (Leucine>Proline) and M>T311 (Methionine>Threonine) were specific only to *G. amoenus* collected from El Faiyum. Compared to other amino acid variants, L>P263 was remarkably less frequent, and it was predicted using PROVEAN database tool to have non-neutral effects.

Conclusion: Amino acid polymorphisms within the cytochrome *b* gene could be assigned to specific geographic locations. They might prove suitable to track accumulated and recent environmental changes as they could represent signs of adaptive evolution.

Keywords: Cytochrome b gene, Amino acid polymorphisms, mtDNA, Biogeographic assignment, *Gerbillus*, Egypt

Background

The Egyptian gerbils of the genus *Gerbillus* are among the more diversified rodent genera of arid and semi-arid regions of the old world. Occupying diverse habitats, these rodents present a useful model for investigating historical environmental changes that affected the arid, old world belt and drove the distribution and morphological diversification in mammals. In Egypt, the distribution and systematics of this genus and related rodent genera have been comprehensively covered (Osborn and Helmy 1980). In this genus, the molecular assessments have brought recent insights in a group of rodents where morphological studies have always dominated (Musser and Carleton 2005; Ndiaye, Chevret, Dobigny, & Granjon, 2016; Ndiaye et al. 2016b).

At the molecular levels, the most widely used population genetic indicator in animals is mitochondrial DNA (Avise et al. 1987). One of the important mitochondrial genes for survival of organisms is the cytochrome *b* (cyt*b*) gene. This gene has proved particularly useful for discerning phylogenetics and taxonomic relations (Castresana 2001; Cook et al. 1999; Irwin et al. 1991; Kuwayama and Ozawa 2000; Kocher et al. 1989; Lau et al. 1998).

The cyt*b* gene encodes a membrane-bound molecule, a central catalytic subunit of ubiquinol cytochrome c reductase (*bc1* complex or complex-III), that is present in the respiratory chain of mitochondria (Howell 1989). Given the exception of protozoans missing mitochondria, all eukaryotes need this class of reduction-oxidation enzyme, and subsequently cyt*b*, for energy metabolism (Hauska et al. 1983; Trumpower 1990; Widget and Cramer 1991). Further studies have evaluated the physiochemical changes which resulted from the molecular evolution of

* Correspondence: mah.khalifa@gmail.com; mahmoud.khalifa@azhar.edu.eg
[1]Department of Zoology, Molecular Biology Lab., Faculty of Science, Al-Azhar University, Nasr City, Cairo, Egypt
Full list of author information is available at the end of the article

cyt*b* functional domains in pocket gophers and cetartiodactyls (McClellan and McCracken 2001).

Although many scientific studies on the assessment of cyt*b* amino acid residues at levels of both structure and function (Esposti et al. 1993) as well as physiochemical changes (McClellan and McCracken 2001) already date back to the end of the twentieth century, it is reasonable for investigation as they become more feasible with freely available annotated collection of mtDNA sequences. Some methods have been employed to discover signs of adaptation, among them identifying local declines in diversity, indicating selective sweeps in different populations of species (Andolfatto 2001; Nielsen 2005). Possibly, recent and established online database tools as SIFT (Ng and Henikoff 2003) and PolyPhen-2 (Adzhubei et al. 2013) that have been implemented for evaluating non-synonymous amino acid changes are also helpful in this process. In consistence with previous tools, prediction and assessment of single amino acid substitution using PROVEAN (Choi et al. 2015) become more attainable.

The present study aims to address the relationships between intra- and interspecies-specific mitochondrial cyt*b* amino acid variants in genus *Gerbillus* from different geographical regions in Egypt. Specific amino acid variants will be assigned to different localities of the Western Desert of Egypt.

Methods

DNA techniques

Analysis was carried out on 20 specimens from different localities in Egypt (Table 1). Specimens were classified based on morphological taxonomic characters described by Harrison and Bates (1991) and Osborn and Helmy (1980). Four species have been identified: *Gerbillus campestris, G. andersoni, G. amoenus,* and *G. gerbillus*. Molecular analyses were done in the Laboratory of Molecular Biology, Zoology Department, Faculty of Science, Al Azhar University, Cairo. Genomic DNA was isolated from femoral muscle, using QIAamp® DNA Mini Kit.

The complete cyt*b* was amplified using Thermocycler GeneAmp 9700, USA. Pairs of primers, L14723 (5′-ACCAAT GAC ATG AAA AAT CAT CGT T-3′) and H15915 (5′-TCT CCATTT CTG GTT TAC AAG AC-3′), as described by Ducroz (1998), were used to target the cytb gene of 1140 bp size. The polymerase chain reaction (PCR) thermal program was set as follows. Initial denaturation at 94 °C for 1 min. Then 35 cycles of denaturation at 94 °C for 1 min, annealing at 62 °C for 1 min, and extension at 72 °C for 1 min. Final extension step at 72 °C for 10 min. PCR amplifications were carried out in 25 μl reaction volume including 0.5 μl of each primer (10 pmole/μl), 12.5 μl 2× PCR Master Mix solution (i-Taq™), 10.5 H_2O and 1 μl of template DNA. PCR products were purified using the QIAquick PCR purification kit and prepared for automated sequencing (3500 Genetic Analyzer, Applied Biosystems) relying on one of the primers utilized for the amplification. The acquired new sequences were all deposited in GenBank under serial accession numbers KX786151 to KX786155 and KX792465 to KX792479 after sequence quality assessments (Table 1).

Genetic data analysis

Sequence and database analysis

MEGA 7.0.14 software (Kumar et al. 2016) was used to align and proofread generated sequences from the sequencer. Conflicting DNA bases within DNA sequences were verified against the associated chromatograms. Accordingly, the cytochrome b sequences were generated for subsequent database analysis.

Basic Local Alignment Search Tool (BLAST)

To check and identify generated sequences, each was blast searched as a query through NCBI (National Center for Biotechnology Information) Blastn tool (www.ncbi.nlm.nih.gov/BLAST/). Sequences with best hits were retrieved and used as outgroups (Table 1) for further comparison to cyt*b* sequences from the current study.

Multiple sequence alignment (MSA)

The nearly complete cyt*b* gene sequences, 20 samples generated in this study were aligned against 24 sequences downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank) of various well characterized *Gerbillus* species as a reference, and *Sekeetamys calurus* was used as an outgroup. Initial analysis, using DnaSP v.510 (Librado and Rozas 2009), of the total 44 cyt*b* gene sequences gave 37 haplotypes represented as Hap1, Hap2, …, and Hap37 that will stand for future alignment.

Multiple sequence alignment was performed for DNA and protein sequences. The Bioinformatics Resource Portal (http://web.expasy.org/translate/) was used to translate the DNA sequence of cyt*b* gene into its respective protein. Functional effect predictions of non-synonymous single nucleotide polymorphisms (nsSNPs) were achieved using PROVEAN (Protein Variation Effect Analyzer) (http://provean.jcvi.org/) (Kumar et al. 2014), which measures the damaging effect of variations in protein sequences (Choi et al. 2012). The prediction is based on the change in the similarity of the sequence to related protein sequences in a MSA by a delta alignment score of the reference and the variant carrying protein sequence with respect to the alignment of homologous sequences (Choi 2012). Variants with score equal to or below − 2.5 are considered as deleterious nsSNP.

Table 1 List of specimens with their GenBank accession numbers of the sequences of cyt*b* (italicized numbers refer to original sequences in this study), geographic origin, haplotype code, and literature source

No.	Species	Accession number	ID	Country	Locality	Geographical coordinates	Haplotype code	Source
1	*G. amoenus*	*KX786153*	M005255	Egypt	Faiyum	29° 34′ N–30° 54′ E	Hap1	Current study
2	*G. amoenus*	*KX792471*	M005256	Egypt	Faiyum	29° 34′ N–30° 54′ E	Hap2	Current study
3	*G. amoenus*	*KX792472*	M005257	Egypt	Faiyum	29° 34′ N–30° 54′ E		Current study
4	*G. amoenus*	*KX792473*	M005258	Egypt	Faiyum	29° 34′ N–30° 54′ E	Hap3	Current study
5	*G. amoenus*	*KX792474*	M005259	Egypt	Faiyum	29° 34′ N–30° 54′ E	Hap4	Current study
6	*G. amoenus*	*KX786154*	M005262	Egypt	Wadi El Natroun	30° 23′ N–30° 22′ E	Hap5	Current study
7	*G. amoenus*	*KX792475*	M005263	Egypt	Wadi El Natroun	30° 23′ N–30° 22′ E		Current study
8	*G. andersoni*	*KX786151*	M005265	Egypt	Baltim-Kafr El Sheikh	31° 34′ N–31° 13′ E	Hap13	Current study
9	*G. andersoni*	*KX792465*	M005266	Egypt	Baltim-Kafr El Sheikh	30° 34′ N–31° 13′ E	Hap14	Current study
10	*G. andersoni*	*KX792466*	M005267	Egypt	Baltim-Kafr El Sheikh	30° 34′ N–31° 13′ E		Current study
11	*G. andersoni*	*KX792467*	M005268	Egypt	Baltim-Kafr El Sheikh	30° 34′ N–31° 13′ E	Hap15	Current study
12	*G. andersoni*	*KX792468*	M005269	Egypt	Baltim-Kafr El Sheikh	30° 34′ N–31° 13′ E	Hap16	Current study
13	*G. andersoni*	*KX792469*	M005274	Egypt	Dabaa- Matruh	30° 59′ N–28° 27′ E	Hap17	Current study
14	*G. andersoni*	*KX792470*	M005275	Egypt	Dabaa- Matruh	30° 59′ N–28° 27′ E		Current study
15	*G. andersoni*	*KX786152*	M005277	Egypt	Dabaa- Matruh	30° 59′ N–28° 27′ E	Hap18	Current study
16	*D. campestris*	*KX786155*	M005278	Egypt	El Zaiytona, Siwa Oasis	29° 09′ N–25° 49′ E	Hap33	Current study
17	*D. campestris*	*KX792478*	M005279	Egypt	El Zaiytona, Siwa Oasis	29° 09′ N–25° 49′ E		Current study
18	*D. campestris*	*KX792479*	M005280	Egypt	El Zaiytona, Siwa Oasis	29° 09′ N–25° 49′ E		Current study
19	*G. gerbillus*	*KX792476*	M005284	Egypt	Dabaa- Matruh	30° 59′ N–28° 27′ E	Hap25	Current study
20	*G. gerbillus*	*KX792477*	M005286	Egypt	Dabaa- Matruh	30° 59′ N–28° 27′ E	Hap26	Current study
21	*G. amoenus*	KM236112	N3009	Niger	Ourou, Air	19° 16′ N–7° 96′ W	Hap6	Ndiaye, Tatard, et al., 2016
22	*G. amoenus*	AJ851270	1997016	Mauritania	Nauakchot	18° 12′ N–16° 02′ W	Hap7	Ndiaye et al. 2013
23	*G. amoenus*	JQ753052	1553	Libya	Sabha	27° 14′ N-14° 24′ E	Hap8	Ndiaye et al. 2013
24	*G. amoenus*	JQ753062	1999032	Mali	Edjerir	18° 12′ N- 1° 24′ W	Hap9	Ndiaye et al. 2013
25	*G. nanus*	JQ753054	Gn2	Israel	Arava	29° 51′ N-35° 03′ E	Hap10	Ndiaye et al. 2013
26	*G. nanus*	JQ753058	Gn6	Israel	Arava	29° 51′ N-35° 03′ E	Hap11	Ndiaye et al. 2013
27	*G. nanus*	JQ753063	1988007	Pakistan	Sind Desert		Hap12	Ndiaye et al. 2013
28	*G. andersoni*	KM236114	NMP48241	Jordon	Wadi Rum	29.61′ N-35.43′ E	Hap19	Ndiaye, Tatard, et al., 2016
29	*G. andersoni*	KF496286	NMP 48323	Libya	Sabkhat karkurah	31° 42′ N- 20° 03′ E	Hap21	Ndiaye et al. 2014
30	*G. andersoni*	KM236113	7044Gand	Israel	Moshav Avshalon	31° 21′ N-34° 34 ′E	Hap20	Ndiaye, Tatard, et al., 2016
31	*G. andersoni*	LN606673	Gand 2	Israel	Western Negev	31° 21′ N-34° 34′ E	Hap22	Ndiaye, Tatard, et al., 2016
32	*G. pyramidum*	JN652808	1557	Libya	Jezero Gabroon	27° 03′ N-14° 26′ E	Hap23	Ndiaye et al. 2012
33	*G. pyramidum*	JN652809	M5982	Mali	Oued Chacheguerene	19° 43′ N-00° 01′ W	Hap24	Ndiaye et al. 2012
34	*G. gerbillus*	LN606679	1999280	Mauritania	Nauakchot	18° 33′ N–15° 96′ W	Hap27	Ndiaye, Tatard, et al., 2016
35	*G. gerbillus*	AJ851269		Mauritania	Nauakchot			Chevret and Dobigny 2005
36	*G. gerbillus*	GU356564	88F	Tunisia	Faouar	33° 16′ N–08° 29′ E	Hap28	Abiadh et al. 2010
37	*G. gerbillus*	JN021403	LG112	Morocco	14 km N of Tarfaya	27° 57 N-12° 46 W	Hap29	Ndiaye et al. 2012
38	*G. gerbillus*	KF496219	M4953	Mali	Vallee Tillemsi	19° 35′ N-0° 183′ W	Hap30	Ndiaye et al. 2014

Table 1 List of specimens with their GenBank accession numbers of the sequences of cyt*b* (italicized numbers refer to original sequences in this study), geographic origin, haplotype code, and literature source *(Continued)*

No.	Species	Accession number	ID	Country	Locality	Geographical coordinates	Haplotype code	Source
39	*G. gerbillus*	LN606680	2002461	Niger	Achegour	19° 01′ N 11° 71′ W	Hap31	Ndiaye, Tatard, et al., 2016
40	*G. gerbillus*	KR089023	CM:113823	Egypt	Giza	29° 70′ N30° 97′ E	Hap32	Alhajeri et al. 2015
41	*D. campestris*	KF496218	1999040	Niger	Air (Agadez)		Hap34	Ndiaye et al. 2014
42	*D. campestris*	AJ851271	1999030	Mali	Edjerir, Adrar des Iforas	18° 2′ N1° 4′ W	Hap35	Chevret and Dobigny 2005
43	*D. campestris*	KR089016	TK-4090	Tunisia	Mournagia Subgovt, Jebel Ain Es Seed		Hap36	Alhajeri et al. 2015
44	*Sekeetamys calurus*	AJ851276		Egypt			Hap37	Chevret and Dobigny 2005

ID refers to the identification number of the samples studied and, when available, to the corresponding museum specimen

Results and discussion

Multiple sequence alignment

Cytb DNA sequence

The alignment of cyt*b* DNA sequences, from the current study and database, became more accurate after using a codon-wise approach which facilitated assuring positions of possible insertions and deletions. Codon-wise sequence alignment of cyt*b* haplotypes from populations of different *Gerbillus* species, together with outgroup haplotypes, exposed most of the segregation sites which were probably enough for speciation events between different taxa.

Out of the 942 nucleotides used in the multiple alignments of cyt*b* haplotypes, 282 variable sites compared to 660 conserved sites. From the variable sites, 38 positions were found as singletons while the remaining 244 variable sites were parsimony informative.

Cytb amino acid sequence

Multiple alignment of amino acid sequences is a standard technique for visualizing the relationships between residues in a collection of evolutionary related proteins. In the current study, multiple alignment of the cyt*b* amino acid residues from *Gerbillus* species gave the chance to deepen the understanding of the evolutionary relationship between some species within the generic name *Gerbillus*.

Out of 313 deduced amino acid residues used in the alignment of cyt*b* haplotypes (Fig. 1), 27 polymorphic sites were compared to 286 conserved sites. Out of these polymorphic sites, 19 sites were parsimony informative sites. Species-specific variations as well as those shared within the genus *Gerbillus* appear to be geographically/habitat related and could possibly be assigned to species biogeographic history. Possibly, these variations can be used to explain evolutionary relationships, as well as the type and effect of natural selection which might act differently in different geographic regions.

DNA-based amino acid polymorphisms

The major species-specific and region-restricted amino acid variants within the cyt*b* sequence of *G. amoenus* and its closely related congeneric species are compiled in Table 2 and Fig. 2. This evaluation of the multiple sequence alignment of cyt*b* amino acid sequences proved helpful for exploring a group of variants, that are species specific, in more details. In addition, more understanding was achieved after using the online PROVEAN (Protein Variation Effect Analyzer) tool for filtering variants that possibly have an impact on the biological function of cyt*b* protein.

In general, among polymorphic sites that were identified from the alignment, 8 amino acid variants V>I2, I>L106, I>V176, L>M222, F>L226, I>L229, L>S237, and I>T295 were specifically portraying the outgroup genus *Sekeetamys calurus* (Hap37), compared to 15 variants V>A30, I>N31, L>F223, L>P263, M>T311, L>F173, A>M203, I>V221, V>M283, V>M30, A>S203, V>L30, F>Y217, I>T229, and V>I283 assigned specifically to species members from genus *Gerbillus* and were enough for separating this genus from its closest outgroup *S. calurus*, while 8 amino acid variants were in common between both genera (Fig. 2, Table 2).

In more details, amino acid variants V>A30 (Valine>Alanine), I>N31 (Isoleucine>Asparagine), L>F223 (Leucine>Phenylalanine) L>P263 (Leucine>Proline), and M>T311 (Methionine>Threonine) showed a restriction to *G. amoenus*, i.e., only 5 amino acid variant positions are specifically characterizing *G. amoenus* members from their closest species. Further, amino acid variants and similar scale of variations that showed a specificity to the other species are assorted in Table 2 and Fig. 2. Therefore, this possibly introduces a usable utility with a respective variation scale for each species. That way, collectively with inclusion of more genes, we demand further investigation to extend this scale of variation to cover more species and genera and develop a utility as a novel trend to probably solve and/or

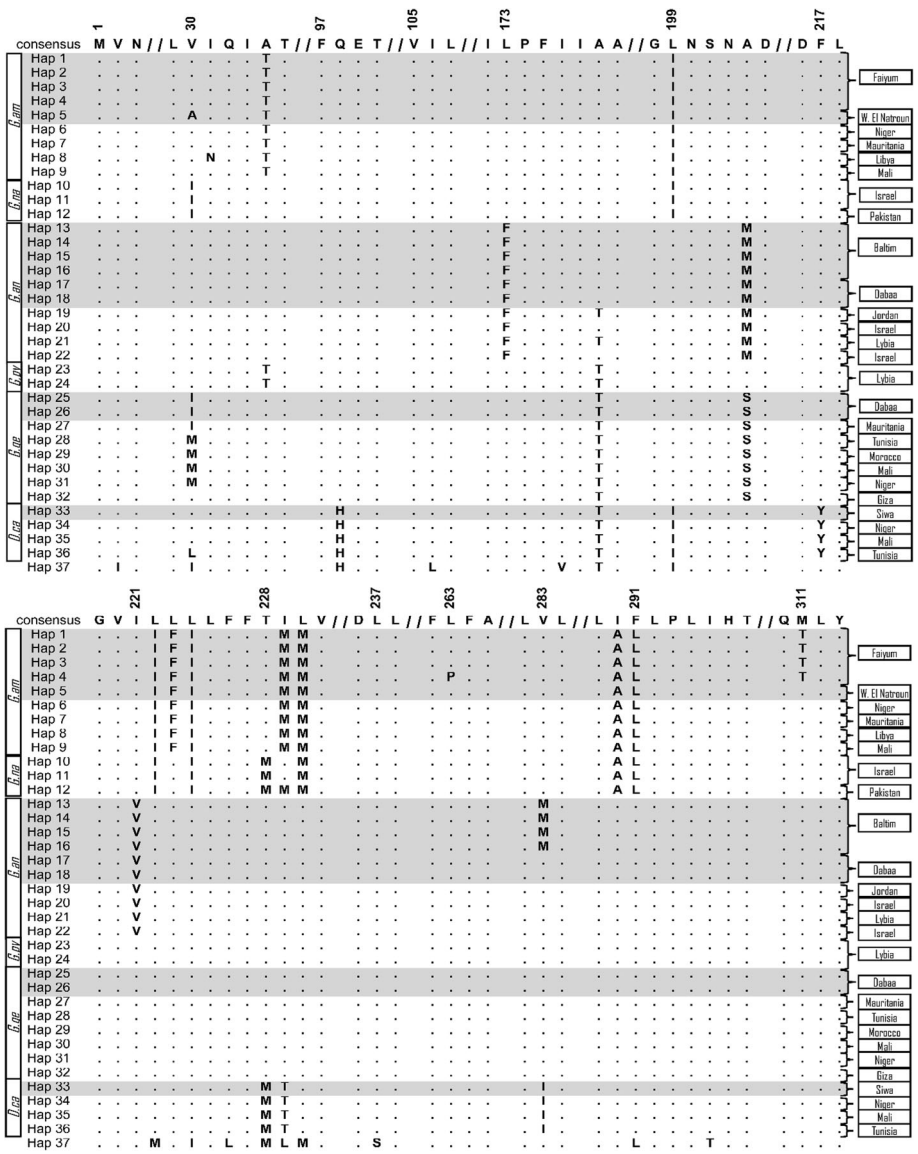

Fig. 1 Multiple sequence alignment of *cytb* protein residues haplotypes. "Hap" indicates haplotype; (.) represent the same amino acid as in the consensus sequence; (/ /) stands for sequences shared between all haplotypes. G.am: *Gerbillus amoenus*; G.na: *Gerbillus nanus*; G.an: *Gerbillus andersoni*; G.py: *Gerbillus pyramidum*; G.ge: *Gerbillus gerbillus*; D.ca: *Dipodillus campestris* (also common as *Gerbillus campestris*). Haplotypes with *gray* background are these from the current study compared to those from the database

reduce recent controversies of molecular and morphologic taxonomy of genus *Gerbillus* reported by Musser and Carleton (2005) and Ndiaye et al. (2016a, 2016b).

In this work, among the evaluation outputs of multiple sequence alignment of cyt*b* protein is the assignment of a group of amino acid variants to specific biogeographic regions. The variants V>A30 in haplotype 5 (Hap5, current study), I>N31 in haplotype 8 (Hap.8, database), L>P263 and M>T311 in haplotypes 1–4 (Hap1–4, current study), V>M283 in haplotypes 13–16 (Hap13–16, current study), and V>L30 in haplotype 36 were respectively assigned to Wadi el Natroun Egypt, Libya, El Faiyum Egypt, Baltim Egypt, and Tunisia (Table 2, Fig. 1). With these different localities, characterized by certain amino acid variants, we can record signs of interaction between environmental changes and wild life of diverse habitat overtime in Egypt in order to better track future changes and answer questions relating to population level.

Database prediction of different amino acid variants using PROVEAN allowed speculation of possible effects on the biological function of cyt*b* protein in different localities. The PROVEAN predicted that the effects of cyt*b* amino acids were mostly of neutral biological function except for amino acid variants L>P263 in *G. amoenus*

Table 2 Intraspecific variation within different species of the *Gerbillus* species

Species	Collection site	Haplotype no.	Amino acid change (species specific)	Nucleotide change	Codon change	PROVEAN		Region restricted variants	Source
						Predicted effect	Scores		
G. amoenus	Wadi el Natroun	Hap5	V>A30	T90C	GTT>GCT	Neutral	− 2.127	V>A30	Current study
	Libya	Hap8	I>N31	T93A	ATT>AAT	Neutral	− 1.930	I>N31	Outgroup
	El Faiyum, Wadi el Natroun Niger, Mauritania Libya, Mali	Hap1–9	L>F223	C668T, A670T	CTA>TTT	Neutral	− 1.553	–	Current study + outgroup
	El Faiyum	Hap4	L>P263	T789C	CTA>CCA	Deleterious	− 4.513	L>P263	Current study
		Hap1–4	M>T311	T933C	ATA>ACA	Neutral	− 0.593	M>T311	
G. andersoni	Jordan Libya	Hap19, Hap21	L>F173	C518T, C520T	CTC>TTT	Neutral	− 1.978	–	Outgroup
	Baltim, Dabaa, Israel	Hap13–18, Hap20, Hap22		C518T	CTC>TTC	Neutral		–	Current study + outgroup
	Baltim, Dabaa, Libya, Jordan Israel	Hap13–22	A>M203	G608A, C609T	GCA>ATA	Neutral	− 1.459	– –	
			I>V221	A662G, T664C	ATT>GTC	Neutral	0.127	–	
	Baltim	Hap13–16	V>M283	G848A	GTA>ATA	Neutral	− 0.015	V>M283	Current study
G. gerbillus	Tunisia, Morocco, Mali, Niger	Hap28–31	V>M30	G89A	GTA>ATA	Neutral	− 1.076	–	Outgroup
	Dabaa, Mauritania, Tunisia, Morocco, Mali, Niger, Giza	Hap25–32	A>S203	G608T	GCA>TCA	Neutral	− 0.364	–	Current study + outgroup
D. campestris	Tunisia	Hap36	V>L30	G89C	GTA>CTA	Neutral	− 0.647	V>L30	Outgroup
	Siwa, Niger, Mali, Tunisia	Hap33–36	F>Y217	T651A	TTC>TAC	Neutral	− 0.715	–	Current study + outgroup
			I>T229	T687C	ATT>ACT	Neutral	1.448	–	
			V>I283	G848A, A850T	GTA>ATT	Neutral	0.183	–	
S. calurus	Egypt	Hap37	V>I2	T5A	GTA>GAA	Neutral	0.034	–	Outgroup
			I>L106	A317C, T319C	ATT>CTC	Neutral	− 1.046	–	
			I>V176	A527G	ATC>GTC	Neutral	− 0.196	–	
			L>M222	C667A	ATC>ATA	Neutral	− 0.708	–	
			F>L226	T677C, C679A	TTC>CTA	Neutral	1.225	–	
			I>L229	A686C	ATT>CTA	Neutral	0.104	–	
			L>S237	T711C	TTA>TCA	Deleterious	− 4.159	–	
			I>T295	T885C, T886A	ATT>ACA	Neutral	− 1.700	–	

collected from El Faiyum (current study) and L>S237 in *S. calurus* from Egypt (database) which were deleterious. These amino acid variants that gave neutral PROVEAN scores are given in Table 2. We are aware that any scores of in silico prediction tools should be taken with caution and might not actually represent the true effect of variants on biological function. The predictive accuracy of these tools was recently studied by Leong et al. (2015) who concluded that they must never be relied on as a final arbiter of pathogenicity but that they should rather be assessed as raising or lowering probabilities. Also, reports by Choi et al. (2012) have underlined that low scores are given to those amino acid residues found in conserved regions or domains while high scores were found in non-conserved regions when searched against the database of related sequences. PROVEAN has a higher specificity score compared to other efficient and commonly used tools like SIFT and PolyPhen-2 (Choi et al. 2012).

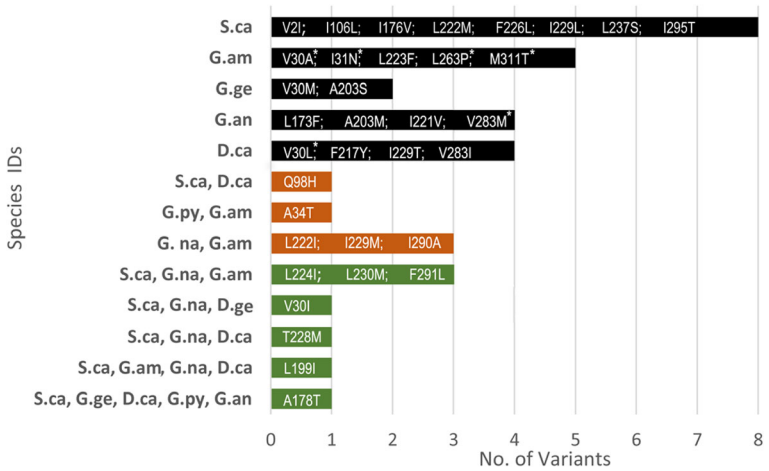

Fig. 2 Cyt*b* amino acid variants within and between species. G & S indicates *Gerbillus* & *Sekeetamys*; na, ca, am, an, cam, ge, & py, represent *nanus*, *calurus*, *amoenus*, *andersoni*, *campestris*, *gerbillus*, & *pyramidum*, respectively. Asterisk (*) is used to indicate variants which are geographically assigned

Interestingly, unlike other amino acid variants that are predicted to give a neutral PROVEAN score, L>P263 (Leucine>Proline), with its relatively small prevalence as a minor allele (Fig. 2, Hap4) in and specific to the El Faiyum depression, is predicted by PROVEAN to carry a deleterious functional effect (Table 2) with a score of –4.513. Based on this predicted effect, it may be expected that *G. amoenus* members carry this "P" allele as a short-lived polymorphism, which does not persist in the population long enough to become fixed and rather become lost overtime due to lack of fitness and further fixation of the alternative "L" allele, possibly because of random genetic drift as similarly explained by Ohta (1992a, 1992b) who studied amino acid polymorphisms in both Drosophila and human mtDNA. However, a relatively similar study (Rand and Kann 1996) on one of the mitochondrial genes gave a different point of view, which suggested that some pressure of adaptive divergence might be acting in conflict with forces removing or rather eliminating variation. Interestingly, this might better describe the running scene of adaptive divergence in El Faiyum Depression. With the low sample size and one out of five gerbils at that location carrying the allele with the non-neutral, and hence possibly deleterious or possibly advantageous effect on the function of the vital cyt*b* protein, we might either have sampled by chance a very rare individual, or the allele frequency might be indeed rather high. The latter scenario would argue against a deleterious effect of L>P263 and thus favor the hypothesis of positive adaptive evolution.

Conclusions

We evaluated the importance of multiple sequence alignment of cytochrome b amino acid sequences of the genus *Gerbillus* from different habitat types in Egypt. This alignment proved useful for assigning specific amino acid variants within and/ or between species and to specific geographic locations of varying environmental conditions. Herewith, the provided data and analysis approach will be a useful starting point for further investigation on the geographic distribution of cyt*b* sequence variants and their possible role in environmental adaptation of Gerbillus spp. in Egypt.

Abbreviations

A: Alanine; BLAST: Basic Local Alignment Search Tool; bp: Base pair; Cyt*b*: Cytochrome *b*; D: Aspartic acid; DnaSP: Deoxyribonucleic acid sequence polymorphism; E: Glutamic acid; F: Phenylalanine; G: Glycine; H: Histidine; Hap: Haplotype; I: Isoleucine; L: Leucine; M: Methionine; MSA: Multiple Sequence Alignments; mtDNA: Mitochondrial deoxyribonucleic acid; N: Asparagine; NCBI: National Center for Biotechnology Information; nsSNP: Non-synonymous single nucleotide polymorphism; P: Proline; PCR: Polymerase chain reaction; PolyPhen2: Polymorphism Phenotyping 2; PROVEAN: Protein Variation Effect Analyzer; Q: Glutamine; S: Serine; SIFT: Sorting Intolerant From Tolerant; SNP: Single nucleotide polymorphism; Spp: Species; T: Threonine; V: Valine; Y: Tyrosine

Acknowledgements

We are grateful to Prof. Dr. Mostafa A. Saleh from the Department of Zoology, Faculty of Science, Al Azhar University, Cairo, Egypt, for his help in collecting the study specimens and for his helpful advice. Dr. Konrad Schmidt, afrigene, Innsbruck, Austria, provided valuable help in editing the revised manuscript, both for language as well as scientific issues.

Funding

Not applicable.

Authors' contributions

MAK is the principle investigator who designed the experiments, led the data analysis, and wrote the paper. MIY is the supervisor who identified and classified the gerbils specimens according to their morphological characteristics and shared in the writing process and revision. AG performed the main practical work. All authors read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Author details
[1]Department of Zoology, Molecular Biology Lab., Faculty of Science, Al-Azhar University, Nasr City, Cairo, Egypt. [2]Department of Zoology, Ecology Lab., Faculty of Science, Al Azhar University, Nasr City, Cairo, Egypt.

Received: 25 May 2017 Accepted: 5 January 2018
Published online: 16 January 2018

References

Abiadh, A., Chetoui, M., Cheniti, T. L., Capanna, E., & Colangelo, P. (2010). Molecular phylogenetics of the genus *Gerbillus* (Rodentia, Gerbillinae): Implications for systematics, taxonomy and chromosomal evolution. *Molecular Phylogenetics and Evolution*, *56*, 513–518.

Adzhubei, I., Jordan, D.M., Sunyaev, S.R., 2013. Predicting functional effect of human missense mutations using PolyPhen-2. Current Protocols in Human Genetics. Chapter 7, Unit7. 20.

Alhajeri, B. H., Hunt, A. J., & Steppan, S. J. (2015). Molecular systematics of gerbils and deomyines (Rodentia: Gerbillinae, Deomyinae) and a test of desert adaptation in the tympanic bulla. *Journal of Zoological Systematics and Evolutionary Research*, *53*, 312–330. https://doi.org/10.1111/jzs.12102.

Andolfatto, P. (2001). Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics & Development*, *11*, 635–641.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., … Saunders, N. C. (1987). Intraspecific phylogeography: The mitochondrial-DNA bridge between population-genetics and systematics. *Annual Review of Ecology and Systematics*, *18*, 489–522.

Castresana, J. (2001). Cytochrome *b* phylogeny and the taxonomy of great apes and mammals. *Molecular Biology and Evolution*, *18*, 465–471.

Chevret, P., & Dobigny, G. (2005). Systematic and evolution of the subfamily Gerbillinae (Mammalia, Rodentia, Muridae). *Molecular Phylogenetics and Evolution*, *35*, 674–688.

Choi, Y. A. (2012). *Fast computation of pairwise sequence alignment scores between a protein and a set of single-locus variants of another protein in proceedings of the ACM conference on bioinformatics, computational biology and biomedicine (BCB 12)*, (pp. 414–417). New York, NY: ACM.

Choi, Y. A., Chan, A. P., & PROVEAN web server (2015). A tool to predict the functional effect of amino acid substitutions and Indels. *Bioinformatics*, *31*(16), 2745–2747.

Choi, Y. A., Sims, G. E., Murphy, S., Miller, J. R., & Chan, A. P. (2012). Predicting the functional effect of amino acid substitutions and Indels. *PLoS One*, *7*, e46688.

Cook, C. E., Wang, Y., & Sensabaugh, G. (1999). A mitochondrial control region and cytochrome *b* phylogeny of sika deer (*Cervus nippon*) and report of tandem repeats in the control region. *Molecular Phylogenetics and Evolution*, *12*, 47–56.

Ducroz, J. F. (1998). *Contribution des approches cytogénétique et moléculaire à l'étude systématique et évolutive des genres de rongeurs Murinae de la «division» Arvicanthis*. France: PhD thesis, Muséum National d'Histoire Naturelle.

Esposti, M. D., De Vries, S., Crimi, M., Ghelli, A., Paternello, T., & Meyer, A. (1993). Mitochondrial cytochrome b: Evolution and structure of the protein. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, *1143*(3), 243–271.

Harrison, D., & Bates, P. (1991). *The mammals of Arabia*, (2nd ed., p. 354). Sevenoaks, Kent: Harrison Zoological Museum Publ.

Hauska, G., Hurt, E., Gabellini, N., & Lockau, W. (1983). Comparative aspects of quinol-cytochrome c/plastocyanin oxidoreductases. *Biochimica et Biophysica Acta*, *726*, 97–133.

Howell, N. (1989). Evolutionary conservation of protein regions in the protonmotive cytochrome *b* and their possible roles in redox catalysis. *Journal of Molecular Evolution*, *29*(2), 157–169.

Irwin, D. M., Kocher, T. D., & Wilson, A. C. (1991). Evolution of the cytochrome *b* gene of mammals. *Journal of Molecular Evolution*, *32*, 128–144.

Kocher, T. D., Thomas, W. K., Meyer, A., Edwards, S. V., Paabo, S., Villablanca, F. X., & Wilson, A. C. (1989). Dynamics of mitochondrial DNA evolution in mammals: Amplification and sequencing with conserved primers. *Proceedings of the National Academy of Sciences*, *86*, 6196–6200.

Kumar, A., Rajendran, V., Sethumadhavan, R., Shukla, P., Tiwari, S., & Purohit, R. (2014). Computational SNP analysis: Current approaches and future prospects. *Cell Biochemistry and Biophysics*, *68*, 233–239.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, *33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw 054.

Kuwayama, R., & Ozawa, T. (2000). Phylogenetic relationships among European red deer, wapiti, and sika deer inferred from mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, *15*, 115–123.

Lau, C. H., Drinkwater, R. D., Yusoff, K., Tan, S. G., Hetzel, D. J., & Barker, J. S. (1998). Genetic diversity of Asian water buffalo (*Bubalus bubalis*): Mitochondrial DNA D-loop and cytochrome *b* sequence variation. *Animal Genetics*, *29*, 253–264.

Leong, I. U., Stuckey, A., Lai, D., Skinner, J. R., & Love, D. R. (2015). Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations. *BMC Medical Genetics*, *16*(1), 1–13. https://doi.org/10.1186/s12881-015-0176-z.

Librado, P., & Rozas, J. (2009). DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*, 1451–1452.

McClellan, D. A., & McCracken, K. G. (2001). Estimating the influence of selection on the variable amino acid sites of the Cytochrome *b* protein functional domains. *Molecular Biology and Evolution*, *18*(6), 917–925.

Musser, G. C., & Carleton, M. D. (2005). Superfamily Muroidea. In D. E. Wilson, & D. M. Reeder (Eds.), *Mammal species of the world: A taxonomic and geographic reference*, (vol. 2, Third ed., pp. 894–1531). Baltimore: Johns Hopkins University Press.

Ndiaye, A., Ba, K., Aniskin, V., Benazzou, T., Chevret, P., Konecny, A., … Granjon, L. (2012). Evolutionary systematics and biogeography of endemic gerbils from Morocco: An integrative taxonomy approach. *Zoologica Scripta*, *41*, 11–28.

Ndiaye, A., Chevret, P., Dobigny, G., & Granjon, L. (2016a). Evolutionary systematics and biogeography of the arid habitat-adapted rodent genus *Gerbillus* (Rodentia, Muridae): A mostly Plio-Pleistocene African history. *Journal of Zoological Systematics and Evolutionary Research*. https://doi.org/10.1111/jzs.12143.

Ndiaye, A., Hima, K., Dobigny, G., Sow, A., Dalecky, A., Ba, K., … Granjon, L. (2014). Integrative study of a poorly known Sahelian rodent species, *Gerbillus nancillus* (Rodentia, Gerbillinae). *Zoologischer Anzeiger*, *253*, 430–439.

Ndiaye, A., Shanas, U., Chevret, P., & Granjon, L. (2013). Molecular variation and chromosomal stability within *Gerbillus nanus* (Rodentia, Gerbillinae): Taxonomic and biogeographic implications. *Mammalia*, *77*, 105–111.

Ndiaye, A., Tatard, C., Stanley, W., & Granjon, L. (2016b). Taxonomic hypotheses regarding the genus *Gerbillus* (Rodentia, Muridae, Gerbillinae) based on molecular analyses of museum specimens. *ZooKeys*, *566*, 145–155.

Ng, P. C., & Henikoff, S. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, *31*(13), 3812–3814.

Nielsen, R. (2005). Molecular signatures of natural selection. *Annual Review of Genetics*, *39*, 197–218.

Ohta, T. (1992a). The nearly neutral theory of molecular evolution. *Annual Review of Ecology and Systematics*, *23*, 263–286.

Ohta, T. (1992b). Theoretical study of near neutrality. II. Effect of subdivided population structure with local extinction and recolonization. *Genetics*, *130*, 917–923.

Osborn, D.J., Helmy, I. (1980). The contemporary land mammals of Egypt (including Sinai). Field. Zool. New series No.5.

Rand, D. M., & Kann, L. M. (1996). Excess of amino acid polymorphism in mitochondrial DNA: Contrasts among genes from drosophila, mice and humans. *Molecular Biology and Evolution*, *13*(6), 735–748.

Trumpower, B. L. (1990). Cytochrome bc_1 complexes of microorganisms. *Microbiological Reviews*, *54*, 101–129.

Widget, W. R., & Cramer, W. A. (1991). *In cell culture and somatic cell genetics of plants*, (vol. 7B, pp. 149–176). New York: Academic Press.