International Research Journal of Biochemistry and Bioinformatics (ISSN-2250-9941) Vol. 1(6) pp. 139-153, July, 2011
Available online http://www.interesjournals.org/IRJBB

Full length Research Paper

Insertion Sequences in mtDNA of Birds and Fish: No Full Length but Some Short Sequences Detected for Which are Obvious Maternal and Signs of Horizontal Transmission

[1, 2*]Yuri PH. Kartavtsev

[1]A.V. Zhirmunsky Institute of Marine Biology, Vladivostok 690041, Russia,
[2]Far Eastern State University, Vladivostok 690095, Russia,
E-mail address: yuri.kartavtsev48@hotmail.com; Tel: +7-4232-311173; fax: +7-4232-310900.

Accepted 28 June, 2011

An analysis of 74 complete mitogenomes in bird and fish species with a focus on the presence of the insertion sequences (IS) in mtDNA revealed that mitogenomes holds no full-length ISs but there are many their sections. In our survey of 74 complete mitogenomes, from 4 to 15 short IS elements per genome were detected with IS-Finder software. Obtained IS elements are probably inverted repeats of real ISs. Both gender-dependent and horizontal transmission routs of these IS segments were obtained among 74 representatives of vertebrate species. The horizontal transmission throughout food chains like fish → bird, although that may be relatively rare for distant lineages, is potentially possible for human consumers of a crude marine food like sushi and sashimi. This important point requires further consideration.

Key words: mtDNA, insertion sequences (IS), transposons, gender-dependent transmission, horizontal transfer, food chains, molecular evolution.

INTRODUCTION

Mitochondrial DNA (mtDNA) is now widely used for phylogenetic surveys in a wide range of taxa. Many sections of mtDNA that code for protein genes or participate in regulation as the control region (CR) are used as genetic markers for investigation of intra- and interspecies diversity. MtDNA is able to accumulate many base substitutions over a long period of time, providing a comparative tool for taxonomic, evolutionary and phylogenetic research (Nei, 1987; Avise, 2000). Molecular phylogenetic approaches use sequence diversities of a single gene, multiple genes or the complete mitochondrial genome (mitogenome). However, if horizontal transmission has an impact on evolutionary processes throughout mobile or transposable elements, then certain reservations are required to maintain scientific precision both in a molecular evolutionary reconstruction and in a mitogenome structure analysis. Certain reservations in conservation policy and genetic security are also necessary if there are signs of horizontal transfer.

Recent genome projects revealed the existence of numerous transposable elements and simpler mobile elements, the insertion sequences (ISs), both in Eukaryotes (Pritham, 2009) and in Prokaryotes (Craig et al., 2002; Siguier et al., 2006a, 2006b; Wagner et al., 2007; Wagner and Chaux, 2008) extending our knowledge on a complexity of life forms, even such relatively simple forms as bacteria. Organization of a typical IS is quite simple. The IS is represented as a modified open reading frame (ORF) in which the terminal inverted repeats (IRs) are obtained that usually labeled as IRL (left inverted repeat) and IRR (right inverted repeat). A single ORF encoding the transposase (Tpase) is normally exist at the entire length of the IS and extending within the IRR sequence. Also within IS short directly repeated sequences (DRs) generated in the target DNA as a consequence of insertion may be find. The Tpase enzyme promoter, p, which is partially localized in IRL, is obtained along with a couple of domains, Domain I vs. II that represent the base pairs

necessary for sequence-specific recognition and binding by the Tpase. More details of IS structure and other features one may find in a relevant literature (Machilion and Chandler, 1998).

In the case of bacterial ISs, only approximately 50 had been analyzed at the nucleotide level in 1989 (Galas and Chandler, 1989), compared to over 3,000 today (Wagner and Chaux, 2008). This is equally true of eukaryotic “ISs” such as *mariner*, derivatives of which have been found in over 240 insect species only and many in fungi, mammals, fish, and plants (Robertson and Lampe, 1995), and related elements such as Tc*1* have also been observed (Radice et al., 1994). This enormous ISs diversity and distribution is astonishing and requires further analysis, classification and understanding the degree of their homology, which stay unresolved in many instances at least between Prokaryotes and Eukaryotes in general and among organelle and nuclear genomes in particular.

This paper will address to some of these issues at the mtDNA level, mostly for nucleotide sequences (sequences, for shortage in a further text). In particular, the detection of IS itself in a mitogenome is the first goal, and the next goal is the consideration of their orthology and transmission pathways. Basically, a description of a routine mitogenome may serve as a source for database and used in a comparative analysis. Such data include a general composition, a primary nucleotide sequence record, and a polypeptide amino acid sequence or CDS translation data. This information is most easily available for instance at GenBank, NCBI (http://www.ncbi.nlm.nih.gov) or at other DNA banks. General believe is that mitogenome, at least in vertebrate animals, has mostly maternal inheritance (Avise, 2000), it is very conservative (Kim et al., 2005; Nagase et al., 2005; Nohara et al., 2005; Kogelnik et al., 2005) and only the CR is flexible. However, some overlaps and indels in the complete mitogenome have been reported (Nohara et al., 2005; Kartavtsev et al., 2007). Also segments of mtDNA were obtained in nuclear DNA (Sorensen and Fleischer 1996; Thomas et al., 1996; Willet-Brozick et al., 2001). At present, there is no universal design to detect IS in a genome. However, suitable software tools for IS search has been recently developed (Siguier et al., 2006b) and one called IS-Finder was applied for bacteria with an additional specialized software, IScan (Wagner et al., 2007). Web-based tool, IS-Finder, was also applied in the current paper to detect IS presence in mtDNA of several dozen bird and fish species. Just to define IS in mitogenome of these and other species is an important task because no one IS observation was described at the time of this manuscript submission.

This paper will be focused mostly to two points: (i) are there any signs of IS presence in mitogenome, and (ii) if ISs are available in mitogenome, by what way they are mainly transmitted: normally by maternal lines or a horizontal pathway is available too?

MATERIALS AND METHODS

In total, 74 complete mtDNA sequences from GenBank of the organelle database were investigated for the IS presence. For this analysis, the IS-Finder software and the bacteria IS Nucleotide Database (http://www-is.biotoul.fr/is.html) were used from July 8 until October 12, 2009. IS-Finder is able to manage with the bacterial database that includes now more than 3,000 ISs. IS-Finder uses basically the Blast algorithm that utilized in GenBank to align IS sequences from the database to target sequence of nucleotides or amino acids. Presently, the analysis of 74 complete mtDNA nucleotide sequences has been basically performed. The output tables of IS-Finder search are exemplified for illustrative purposes with only 3 species (Table 1). For simplicity each IS family was considered as a single variable or a duplicated variable; however, family members are actually represented by non-identical, frequently unique for each genome ISs as exemplified in Table 1. For our purpose creation of such complex variables is quite suitable, because they give a comparative tool, but we have to keep this complexity in mind.

The list of analyzed bird and fish species and the accession numbers of their complete nucleotide sequences at mtDNA used for IS search are given in Table 2. Complete mtDNA sequences in a Fasta-format were applied to IS-Finder in a line and the BLASTN program unit was run. Default align parameters were basically used and Blosum62 reference matrix was set for calculations when BLASTP program used for polypeptide comparison. The increased gap penalties, both pairwise and multiple, up to 5.0 were also applied for some alignment sessions. Such sessions gave no visible differences in the IS list detected when the same sequences were tested. The output tables of this search were used to establish an MS-Exel database for further statistical analysis. This database included IS’s align scores (S-score, bits and E-value) as variables’ estimated scores for listed IS families. Database comprised in total by 64 variables including two grouping variables, G1 and G2 (Table 2-3). The G1 values were set as an integer value 1 for all bird records in the rows and as 2 for fish (Table 2-3). The G2 (Table 2-3, column three) was defined as follows: for birds score 1 denotes water realm species that can easily have fish as food, score 2 denotes continent realm species that cannot easily have fish as food; for fish score 1 denotes species that can easily be food for birds, score 2 denotes species that cannot easily be bird’s food. Some variables were represented actually by two IS family members. For such cases, extra variables were added. These kinds of variables were created to meet somehow variation in IS number when an IS was represented by 2 or more members (see also explanation below). Cases (rows) in this database were represented by 74 mentioned complete mtDNA sequences of different birds and fish

Table 1: An example of data source: A complete mtDNA nucleotide sequence of the goose, *Anser albifrons*, the crown, *Corvus frugilegus* and the torrent catfish, *Liobagrus obesus* that are able to align specifically with the ISs in the bacterial database by the IS-FINDER software

Sequences producing significant alignments*	IS Family	Group	Origin	Score (bits)	E(value)
Anser albifrons **(AF363031 Aves 1)**					
IS*662*	IS*1182*		*Bacillus halodurans C-125*	36	0.91
IS*Fnu1*	IS*1182*		*Fusobacterium nucleatum subsp. nucleatum*	34	3.6
IS*C1078*	IS*630*		*Sulfolobus solfataricus*	34	3.6
IS*231H*	IS*4*	IS*231*	*Bacillus thuringiensis subsp. darmstadiensis 73-E-10-2*	34	3.6
IS*231G*	IS*4*	IS*231*	*Bacillus thuringiensis subsp. darmstadiensis 73-E-10-2*	34	3.6
IS*1221H*	IS*3*	IS*150*	*Mycoplasma hyopneumoniae*	34	3.6
Corvus frugilegus **(CFR18522, Aves 2)**					
IS*Rle7*	IS*6*		*Rhizobium leguminosarum*	38	0.23
IS*C1078*	IS*630*		*Sulfolobus solfataricus*	38	0.23
IS*Dre4*	IS*1380*		*Desulfotomaculum reducens*	36	0.92
IS*Ppr10*	IS*As1*		*Photobacterium profundum*	36	0.92
IS*Mba11*	IS*1634*		*Methanosarcina barkeri str. fusaro*	36	0.92
IS*Mma18*	IS*1634*		*Methanosarcina mazei Go1*	36	0.92
IS*Mac6*	IS*1634*		*Methanosarcina acetivorans*	36	0.92
IS*Bce1*	IS*1182*		*Bacillus cereus*	34	3.7
IS*Tde2*	IS*30*		*Treponema denticola*	34	3.7
IS*Pma1*	IS*1595*	IS*1595*	*Photobacterium mandapamensis*	34	3.7
IS*Plu9*	IS*4*	IS*50*	*Photorhabdus luminescens subsp. laumondii*	34	3.7
IS*Psp3*	IS*5*	IS*5*	*Pseudomonas sp. KKS102*	34	3.7
IS*Ftu1*	IS*630*	-	*Francisella tularensis subsp. tularensis*	34	3.7
IS*200S*	IS*200*/IS*605*	IS*200*	*Streptococcus pneumoniae type 23*	34	3.7
IS*1414*	IS*256*	-	*Escherichia coli enterotoxigenic 27D*	34	3.7
Liobagrus obesus **(DQ321752)**					
ISTel2	IS*200/IS605*		*Thermosynechococcus elongatus*	36	1.1
ISMgi1	IS*1380*		*Mycobacterium gilvum*	36	1.1
ISSba20	IS*21*		*Shewanella baltica*	36	1.1
ISAar32	IS*1380*		*Arthrobacter arilaitensis*	34	4.3
ISAar10	IS*1380*		*Arthrobacter arilaitensis*	34	4.3
ISBce1	IS*1182*		*Bacillus cereus*	34	4.3
ISCth14	IS*110*		*Clostridium thermocellum*	34	4.3
ISCb2	IS*4*	IS*231*	*Clostridium beijerincki*	34	4.3
ISMba14	ISH*3*		*Methanosarcina barkeri str. fusaro*	34	4.3
ISMac7	IS*200*/IS*605*		*Methanosarcina acetivorans*	34	4.3
IS231K	IS*4*	IS*231*	*Bacillus cereus*	34	4.3

Note. *) Significant alignments in IS-Finder are those which matches create E-values =< 10.

Table 2. Analyzed bird and fish species and accession numbers of 74 complete mtDNA nucleotide sequences used in the paper with information on IS number and IS family studied

No	Species analyzed, mtDNA Sequence Accession Number	Species Group (G2)	IS Num-ber in Mito-genome	IS Family*)
Bird Species (G1 = 1)				
1	*Alectura lathami, AY346092*	2	7	IS*110*, IS*3*, IS*3*, IS*H3*, IS*4*, IS*5*, IS*630*
2	*Anomalopteryx didiformis, AF338714*	2	4	IS*1*, IS*5*, IS*607*, IS*NCY*
3	*Apteryx haastii, AF338708*	2	6	IS*4*, IS*5*, IS*21*, IS*H3*, IS*200*/IS*605*, IS*256*
4	*Anser albifrons, AF363031*	1	6	IS*1182*, IS*1182*, IS*3*, IS*4*, IS*4*, IS*630*
5	*Anseranas semipalmata, AY309455*	1	8	IS*4*, IS*5*, IS*3*, IS*607*, IS*5*, IS*630*, IS*481*, IS*5*
6	*Arenaria interpres, AY074885*	1	8	IS*H3*, IS*4*, IS*4*, IS*200*/IS*605*, IS*256*, IS*91*, IS*As1*, IS*1182*
7	*Aythya americana, AF090337*	1	9	IS*5*, IS*6*, IS*630*, IS*21*, IS*200*/IS*605*, IS*1634*, IS*1380*, IS*91*, IS*NCY*
8	*Branta canadensis, DQ019124*	1	7	IS*4*, IS*NCY*, IS*1380*, IS*As1*, IS*5*, IS*630*, IS*91*
9	*Buteo buteo, AF380305*	2	11	IS*630*, IS*91*, IS*3*, IS*1380*, IS*200*/IS*605*, IS*5*, IS*3*, IS*4*, IS*4*, IS*1634*, IS*110*
10	*Casuarius casuarius, AF338713*	2	12	IS*1380*, IS*4*, IS*4*, IS*4*, IS*481*, IS*1380*, IS*1182*, IS*200*/IS*605*, IS*1595*, IS*4*, IS*200*/IS*605*, IS*1*
11	*Cathartes aura, AY463690*	2	7	IS*1182*, IS*4*, IS*4*, IS*4*, IS*200*/IS*605*, IS*630*, IS*L3*
12	*Ciconia boyciana, AB026193*	1	8	IS*5*, IS*4*, IS*1380*, IS*5*, IS*4*, IS*1634*, IS*630*, IS*256*
13	*Ciconia ciconia, AB026818*	1	9	IS*5*, IS*4*, IS*1182*, IS*1380*, IS*1182*, IS*As1*, IS*5*, IS*1634*, IS*630*
14	*Corvus frugilegus, CFR18522*	2	15	IS*6*, IS*630*, IS*1380*, IS*As1*, IS*1634*, IS*1634*, IS*1634*, IS*1182*, IS*30*, IS*1595*, IS*4*, IS*5*, IS*630*, IS*200*/IS*605*, IS*256*
15	*Coturnix chinensis, AB073301*	2	5	IS*4*, IS*4*, IS*110*, IS*607*, IS*630*
16	*Coturnix japonica, AP003195*	2	8	IS*607*, IS*NCY*, IS*200*/IS*605*, IS*L3*, IS*630*, IS*6*, IS*91*, IS*NCY*
17	*Cygnus columbianus, DQ083161*	1	6	IS*5*, IS*1380*, IS*5*, IS*630*, IS*630*, IS*3*
18	*Dinornis giganteus, AY016013*	2	7	IS*NCY*, IS*110*, IS*4*, IS*256*, Tn*3*, Tn*3*, IS*3*
19	*Diomedea melanophris, AY158677*	1	12	IS*256*, IS*630*, IS*L3*, IS*1380*, IS*982*, IS*NCY*, IS*256*, IS*1380*, IS*21*, IS*91*, IS*5*, IS*L3*
20	*Dromaius novaehollandiae, AF338711*	2	5	IS*3*, IS*607*, IS*110*, IS*3*, IS*607*
21	*Emeus crassus, AY016015*	2	6	IS*NCY*, IS*1*, IS*982*, IS*5*, IS*3*, IS*3*
22	*Eudromia elegans, AF338710*	2	11	IS*3*, IS*5*, IS*L3*, IS*1634*, IS*110*, IS*H3*, IS*21*, IS*L3*, IS*5*, IS*200*/IS*605*, IS*200*/IS*605*
23	*Eudyptula minor, AF362763*	1	5	IS*4*, IS*1182*, IS*1634*, IS*110*, IS*630*
24	*Falco peregrinus, AF090338*	2	7	IS*1182*, IS*6*, IS*1182*, IS*1380*, IS*5*, IS*200*/IS*605*, IS*5*
25	*Gallus gallus, MIGGX*	2	7	IS*5*, IS*1*, IS*1380*, IS*630*, IS*256*, IS*110*, IS*1634*
26	*Gallus gallus bankiva, AP003323*	2	7	IS*5*, IS*1*, IS*1380*, IS*256*, IS*110*, IS*1634*, IS*630*
27	*Gallus gallus gallus, AP003322*	2	6	IS*5*, IS*1*, IS*1380*, IS*256*, IS*110*, IS*1634*
28	*Gallus gallus, AP003321*	2	6	IS*5*, IS*1*, IS*1380*, IS*256*, IS*110*, IS*1634*
29	*Gallus sonneratii, AP006741*	2	9	IS5, IS1, IS1380, IS256, IS21, IS701, IS1634, IS5, IS1, IS1380, IS256, IS1634, IS630, IS4
30	*Gallus varius, AP003324*	2	6	IS5, IS1380, IS110, IS630, IS30, IS4
31	*Gavia stellata, AY293618*	1	13	IS3, IS30, IS1182, IS701, IS1182, IS1182, ISAs1, IS91, IS5, IS5, IS1, IS1, IS630
32	*Haematopus ater, AY074886*	1	9	IS5, IS4, IS66, IS5, IS5, IS200/IS605, IS607, IS630, ISL3

Table 2, continue 1

33	*Larus dominicanus, AY293619*	1	4	S*4*, IS*110*, IS*110*, IS*3*
34	*Ninox novaeseelandiae, AY309457*	2	9	IS*630*, IS*4*, IS*66*, IS*66*, IS*6*, IS*66*, IS*200*/IS*605*, IS*L3*, IS*5*
35	*Numida meleagris, AP005595*	2	6	IS*66*, IS*21*, IS*L3*, IS*630*, IS*256*, IS*256*
36	*Strigops habroptilus, AY309456*	2	9	IS*66*, IS*NCY*, IS*1*, IS*66*, IS*200*/IS*605*, IS*630*, IS*6*, IS*4*, IS*3*
37	*Pterodroma brevirostris, AY158678*	1	9	IS*1182*, IS*3*, IS*1380*, IS*66*, IS*21*, IS*L3*, IS*4*, IS*630*, IS*110*
38	*Smithornis sharpei, AF090340*	2	12	IS*607*, IS*6*, IS*481*, IS*3*, IS*481*, IS*1634*, IS*630*, IS*5*, IS*5*, IS*5*, IS*5*, IS*5*, IS*5*
39	*Spizaetus alboniger, AP008239*	2	9	IS*4*, IS*607*, IS*NCY*, IS*As1*, IS*5*, IS*L3*, IS*630*, IS*630*, IS*L3*
40	*Spizaetus nipalensis, AP008238*	2	10	IS*607*, IS*1380*, IS*NCY*, IS*As1*, IS*256*, IS*5*, IS*L3*, IS*630*, IS*L3*, IS*3*
41	*Struthio camelus, AF338715*	2	4	IS*1380*, IS*66*, IS*5*, IS*1634*
42	*Tinamus major, AF338707*	2	7	IS*6*, IS*1380*, IS*110*, IS*4*, IS*3*, IS*3*, IS*110*
43	*Vidua chalybeata, AF090341*	2	7	IS*3*, IS*NCY*, IS*110*, IS*982*, IS*630*, IS*200*/IS*605*, IS*200*/IS*605*
44	*Amia calva, AB042952*	1	5	IS*3*, IS*6*, IS*1380*, IS*110*, IS*As1*
45	*Acanthogobius hasta, AY486321*	1	11	IS*30*, IS*3*, IS*1182*, IS*3*, IS*3*, IS*1595*, IS*4*, IS*1595*, IS*630*, IS*As1*, IS*701*
Fish Species (G1 = 2)				
46	*Acipenser dabryanus, AY510085*	1	8	IS*21*, IS*30*, IS*1182*, IS*110*, IS*3*, IS*H3*, IS*256*, IS*5*
47	*Acipenser stellatus, AJ585050*	1	9	IS*21*, IS*30*, IS*3*, IS*1595*, IS*607*, IS*H3*, IS*110*, IS*256*, IS*5*
48	*Acipenser transmontanus, AB042837*	1	8	IS*21*, IS*110*, IS*30*, IS*110*, IS*3*, IS*H3*, IS*256*, IS*5*
49	*Albula glossodonta, AP002973*	1	8	IS*30*, IS*4*, IS*4*, IS*5*, IS*4*, IS*256*, IS*As1*, IS*5*
50	*Aldrovandia affinis, AP00297*	2	9	IS*1595*, IS*4*, IS*3*, IS*3*, IS*4*, IS*256*, IS*110*, IS*5*, IS*NCY*
51	*Allocyttus niger, AP004435*	2	11	IS*4*, IS*1380*, IS*200*/IS*605*, IS*256*, IS*256*, IS*5*, IS*630*, IS*3*, IS*200*/IS*605*, IS*200*/IS*605*, IS*256*
52	*Cobitis sinensis, AY526868*	1	8	IS*NCY*, IS*NCY*, IS*3*, IS*NCY*, IS*H3*, IS*3*, IS*30*, IS*256*
53	*Cobitis striata, AB054125*	1	9	IS*3*, IS*1380*, IS*3*, IS*3*, IS*1182*, IS*As1*, IS*3*, IS*30*, IS*4*
54	*Cololabis saira, AP002932*	1	9	IS*21*, IS*4*, IS*3*, IS*4*, IS*4*, IS*1595*, IS*200*/IS*605*, IS*256*, IS*5*
55	*Conger myriaster, AB038381*	2	12	IS*1380*, IS*4*, IS*5*, IS*5*, IS*5*, IS*1380*, IS*1380*, IS*3*, IS*66*, IS*3*, IS*200*/IS*605*, IS*1634*
56	*Coregonus lavaretus, AB034824*	1	14	IS*L3*, IS*256*, IS*As1*, IS*1595*, IS*30*, IS*3*, IS*200*/IS*605*, IS*21*, IS*110*, IS*630*, IS*4*, IS*NCY*, IS*256*, IS*4*
57	*Coreoleuciscus splendidus, DQ347951*	1	10	IS*1380*, IS*30*, IS*1380*, IS*256*, IS*3*, IS*3*, IS*200*/IS*605*, IS*1634*, IS*1634*, IS*4*
58	*Oryzias latipes, AP004421*	1	8	IS*1182*, IS*66*, IS*L3*, IS*3*, IS*H3*, IS*110*, IS*607*, IS*6*
59	*Osteoglossum bicirrhosum, AB043025*	1	7	IS982, IS1595, IS4, IS1380, IS3, ISL3, IS630
60	*Ostichthys japonicus, AP004431*	1	7	IS4, IS4, IS1380, IS607, IS4, ISL3, IS256
61	*Pagrus auriga, AB124801*	2	10	IS1380, IS607, ISAs1, IS1380, IS1182, IS4, IS1595, IS630, IS256, IS256
62	*Pagrus major, AP002949*	2	15	IS5, IS21, IS4, ISH3, IS3, IS1380, ISL3, IS1595, IS1595, IS4, IS4, IS5, Tn3, IS256, IS256

Table 2, continue 2

63	*Pangasianodon gigas, AY762971*	1	9	IS*As1*, IS*5*, IS*5*, IS*1380*, IS*1380*, IS*1182*, IS*5*, IS*1380*, IS*3*
64	*Pantodon buchholzi, AB043068*	1	11	IS*66*, IS*4*, IS*1*, IS*1595*, IS*1595*, IS*1595*, IS*5*, IS*701*, IS*630*, IS*982*, IS*256*
65	*Paralichthys olivaceus, AB028664*	1	9	IS*630*, IS*NCY*, IS*1380*, IS*91*, IS*1380*, IS*21*, IS*1*, IS*1634*, IS*4*
66	*Parazen pacificus, AP004433*	2	9	IS*4*, IS*110*, IS*66*, IS*481*, IS*1595*, IS*1380*, IS*21*, IS*3*, IS*256*
67	*Percopsis transmontana, AP002928*	1	5	IS*630*, IS*110*, IS*4*, IS*L3*, IS*256*
68	*Petroscirtes breviceps, AP004450*	2	13	IS*4*, IS*30*, IS*L3*, IS*982*, IS*982*, IS*982*, IS*982*, IS*110*, IS*5*, IS*110*, IS*481*, IS*As1*, IS*1595*
69	*Platytroctes apus, AP004107*	2	14	IS*3*, IS*5*, IS*21*, IS*1595*, IS*1595*, IS*30*, IS*1380*, IS*3*, IS*5*, IS*L3*, IS*3*, IS*256*, IS*30*, IS*30*
70	*Platichthys bicoloratus, AP002951*	1	8	IS*91*, IS*4*, IS*481*, IS*5*, IS*630*, IS*4*, IS*3*, IS*630*
71	*Phenacogrammus interruptus, AB054129*	1	13	IS*4*, IS*256*, IS*As1*, IS*110*, IS*4*, IS*5*, IS*NCY*, IS*5*, IS*607*, IS*200*/IS*605*, IS*200*/IS*605*, IS*5*, IS*5*
72	*Physiculus japonicus, AP004409*	2	11	IS*1380*, IS256, IS*200*/IS*605*, IS*110*, IS*6*, IS630, IS*4*, IS*256*, IS*91*, IS*256*, IS*5*
73	*Plecoglossus altivelis, AB047553*	1	6	IS*1380*, IS*L3*, IS*30*, IS*110*, IS*701*, IS*200*/IS*605*
74	*Polymixia japonica, AB034826*	2	12	IS*1380*, IS*1380*, IS*1182*, IS*1182*, IS*3*, IS*607*, IS*3*, IS*6*, IS*1595*, IS*5*, IS*256*, IS*4*

Note. *) IS order is shown as it is in IS-Finder output tables, i.e. along with S increase and E score decrease. IS Tn3 are not included in the frequency analysis of IS as being present only in three cases (#18, #18, #62). Grouping variables, G1 and G2 denote different grouping pattern under statistical analysis. G1 scores were set to the value 1 for birds and to the value 2 for fish. Order numbers 1-45 list bird species and order numbers 46-74 listed fish species, correspondingly. The G2 was defined as follows: For birds, 1 denotes water realm species that can easily have fish as a food, 2 denotes continent realm species that cannot easily be fish eaters. For fish, 1 denotes species that can easily be a food for birds, 2 denotes species that cannot easily be bird's food.

species: 45 were birds and 29 were fish (Table 2-3).

As noted, some IS family members were met twice or more times in the same mitogenome (Table 2, e.g. IS*3*, IS*4*, IS*110*, IS*1182*, etc.). For such cases, variables are denoted in mode, IS*110*-S and IS*110*-S1 or IS*110*-E and IS*110*-E1, etc. The consequent score, e.g. for IS*110*-S, was taken as belonging to the first copy in the list of IS members and IS*110*-S1, as belonging to the second IS member (all the rest were ignored). The cases in this database are represented as mentioned by 74 mtDNA sequences. There were many blank spaces in each row in the database, because many certain IS are rare and not present in each mitogenome. These blank spaces were treated as zero (align score = 0) in further calculations. The whole database is too big to put it in the paper even if shorten for statistical analysis (Statsoft data table is available upon request).

Beyond IS-Finder, the statistical analysis was performed using STATISTICA 6.0 (StatSoft, 2001) software package. From this package, the basic module for calculation of the mean and variance parameters, as well as those for canonical analysis and parametric analysis of variance (ANOVA, and multi-dimensional version, MANOVA) and Kruskall-Wallis nonparametric ANOVA were employed.

RESULTS AND DISCUSSION

First, let us consider whether there are any signs of IS presence in mtDNA nucleotide sequences. The names of some peculiar ISs, their bacterial host source, and alignment scores in a certain complete mitogenome (e.g. the goose, *Anser albifrons*, the crown, *Corvus frugilegus,* and the torrent catfish, *Liobagrus obesus*) are exemplified earlier (Table 1). From 4 up to 15 representatives of the known IS families were detected in 74 mitogenomes, their full list is given in Table 2. Some of these families were repeated (Table 2, last column) but exactly the same ISs were rarely detected in a single mitogenome. In general, frequency distribution of IS in mtDNA analyzed is very stochastic (Figure 1). Typically, as exemplified for simplicity for 10 complete mtDNA in top histogram in Figure 1, an individual IS was presented in a particular mitogenome by either one copy or not present at all. Such distribution is quite complicated for a statistical

Table 3. Sample of data table that represents the grouping variables (G1 and G2) and the alignment estimation scores of IS family representatives made by IS-Finder from complete mtDNA nucleotide sequences of 74 representatives of bird and fish species

Species	Grouping Variable		Sequence Alignment scores of IS Family Representatives*: S-Score/E-value							
	G1	G2	IS*110*		IS*3*		IS*30*		IS*5*	
Alectura lathami AY346092	1	2	38	0.23	36	0.91	0	0	34	3.6
Anomalopteryx didiformis AF338714	1	2	0	0	0	0	0	0	34	3.6
Apteryx haastii AF338708	1	2	0	0	0	0	0	0	0	0
Anser albifrons AF363031	1	1	0	0	34	3.6	0	0	0	0
Anseranas semipalmata AY309455	1	1	0	0	0	0	0	0	34	3.6
Arenaria interpres AY074885	1	1	0	0	0	0	0	0	0	0
Aythya americana AF090337	1	1	0	0	0	0	0	0	34	3.6
Branta canadensis DQ019124	1	1	0	0	0	0	0	0	34	3.6
Buteo buteo AF380305	1	2	0	0	0	0	0	0	0	0
Casuarius casuarius AF338713	1	2	0	0	0	0	0	0	0	0
Cathartes aura AY463690	1	2	0	0	0	0	0	0	0	0
Ciconia boyciana AB026193	1	1	0	0	0	0	0	0	34	3.8
Ciconia ciconia AB026818	1	1	0	0	0	0	0	0	34	3.7
Corvus frugilegus CFR18522	1	2	0	0	0	0	34	3.7	34	3.7
Coturnix chinensis AB073301	1	2	36	0.91	0	0	0	0	0	0
Coturnix japonica AP003195	1	2	0	0	0	0	0	0	0	0
Cygnus columbianus DQ083161	1	1	0	0	34	3.6	0	0	36	0.91
Dinornis giganteus AY016013	1	2	34	3.7	34	3.7	0	0	0	0
Diomedea melanophris AY158677	1	1	0	0	0	0	0	0	34	3.7
Dromaius novaehollandiae AF338711	1	2	34	3.6	38	0.23	0	0	0	0
Emeus crassus AY016015	1	2	0	0	34	3.7	0	0	34	3.7
Eudromia elegans AF338710	1	2	34	3.9	0	0	0	0	36	1
Eudyptula minor AF362763	1	1	34	3.8	0	0	0	0	0	0
Falco peregrinus AF090338	1	2	34	3.6	0	0	0	0	34	3.9
Gallus gallus MIGGX	1	2	0	0	0	0	0	0	38	0.23
Gallus gallus bankiva AP003323	1	2	0	0	0	0	0	0	38	0.23
Gallus gallus gallus AP003322	1	2	34	3.6	0	0	0	0	38	0.23
Gallus gallus AP003321	1	2	34	3.6	0	0	0	0	38	0.23
Gallus sonneratii AP006741	1	2	0	0	0	0	0	0	38	0.23
Gallus varius AP003324	1	2	34	3.6	0	0	34	3.6	38	0.23
Gavia stellata AY293618	1	1	0	0	36	0.96	36	0.96	34	3.8
Haematopus ater AY074886	1	1	0	0	0	0	0	0	38	0.23
Larus dominicanus AY293619	1	1	34	3.6	34	3.6	0	0	0	0
Ninox novaeseelandiae AY309457	1	2	0	0	0	0	0	0	34	3.5
Numida meleagris AP005595	1	2	0	0	0	0	0	0	0	0
Strigops habroptilus AY309456	1	2	0	0	0	0	0	0	0	0
Pterodroma brevirostris AY158678	1	1	34	3.6	36	0.9	0	0	0	0
Smithornis sharpei AF090340	1	2	0	0	34	3.8	0	0	34	3.8
Spizaetus alboniger AP008239	1	2	0	0	0	0	0	0	34	3.9
Spizaetus nipalensis AP008238	1	2	0	0	34	3.8	0	0	34	3.8
Struthio camelus AF338715	1	2	0	0	0	0	0	0	34	3.6
Tinamus major AF338707	1	2	36	0.92	34	3.6	0	0	0	0
Vidua chalybeata AF090341	1	2	34	3.7	36	0.93	0	0	0	0
Amia calva AB042952	2	1	34	3.5	36	0.88	0	0	0	0
Acanthogobius hasta AY486321	2	1	0	0	38	0.23	38	0.23	0	0
Acipenser dabryanus AY510085	2	1	34	3.5	34	3.5	36	0.9	34	3.5
Acipenser stellatus AJ585050	2	1	34	3.5	34	3.5	36	0.9	34	3.5
Acipenser transmontanus AB042837	2	1	36	0.91	34	3.6	36	0.91	34	3.6
Albula glossodonta AP002973	2	1	0	0	0	0	36	0.89	34	3.5
Aldrovandia affinis AP00297	2	2	34	3.6	34	3.6	0	0	34	3.6
Allocyttus niger AP004435	2	2	0	0	34	3.6	0	0	34	3.6
Cobitis sinensis AY526868	2	1	0	0	36	0.9	34	3.6	0	0
Cobitis striata AB054125	2	1	0	0	38	0.23	34	3.6	0	0

Table 3, continue 1

Cololabis saira AP002932	2	1	0	0	34	3.6	0	0	34	3.6
Conger myriaster AB038381	2	2	0	0	34	4	0	0	36	1
Coregonus lavaretus AB034824	2	1	34	3.6	36	0.91	36	0.91	0	0
Coreoleuciscus splendidus DQ347951	2	1	0	0	34	3.6	36	0.9	0	0
Oryzias latipes AP004421	2	1	34	3.6	34	3.6	0	0	0	0
Osteoglossum bicirrhosum AB043025	2	1	0	0	34	3.5	0	0	0	0
Ostichthys japonicus AP004431	2	1	0	0	0	0	0	0	0	0
Pagrus auriga AB124801	2	2	0	0	0	0	0	0	0	0
Pagrus major AP002949	2	2	0	0	34	3.7	0	0	38	0.24
Pangasianodon gigas AY762971	2	1	0	0	34	3.6	0	0	36	0.9
Pantodon buchholzi AB043068	2	1	34	3.4	0	0	34	3.4	34	3.4
Paralichthys olivaceus AB028664	2	1	0	0	0	0	0	0	0	0
Parazen pacificus AP004433	2	2	36	0.92	34	3.6	0	0	0	0
Percopsis transmontana AP002928	2	1	34	3.5	0	0	0	0	0	0
Petroscirtes breviceps AP004450	2	2	34	3.6	0	0	36	0.91	34	3.6
Platytroctes apus AP004107	2	2	0	0	38	0.23	36	0.91	36	0.91
Platichthys bicoloratus AP002951	2	1	0	0	0	0	0	0	34	3.4
Phenacogrammus interruptus AB054129	2	1	34	3.6	0	0	0	0	34	3.6
Physiculus japonicus AP004409	2	2	34	3.7	0	0	0	0	34	3.7
Plecoglossus altivelis AB047553	2	1	34	3.6	0	0	38	0.23	0	0
Polymixia japonica AB034826	2	2	0	0	34	3.6	0	0	34	3.6

Note. *) The whole list of ISs is shown in the Table 2. Values for the G1 and G2 variables are given repeatedly as in Table 2 for convenience.

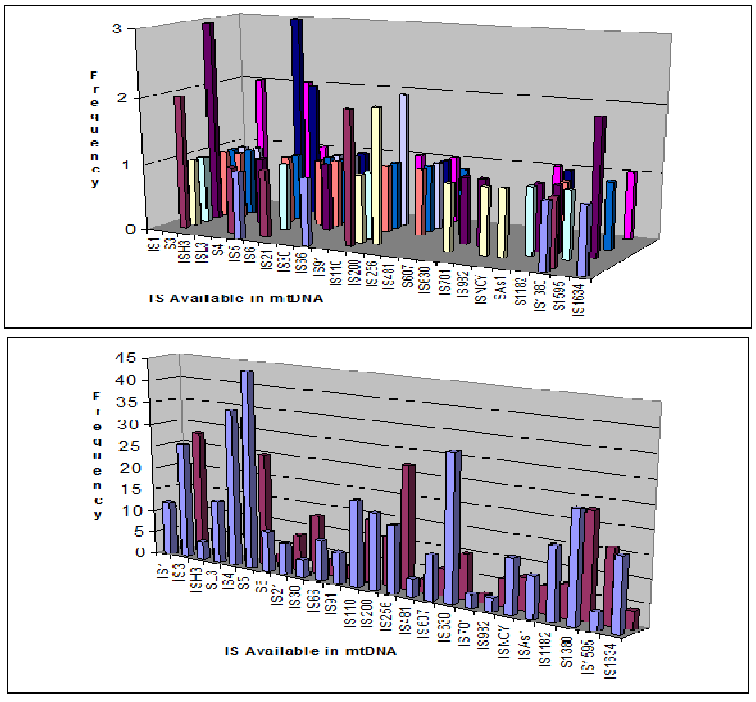

Figure 1. IS frequency distribution for bird and fish species.
The top histogram presents the frequency distribution in bird (5 front rows; #41-45 from Table 1) and fish species (5 back rows; #46-50 from Table 1). The bottom histogram summarizes the frequency distribution for 45 bird (front row) and 29 fish species (back row). IS Tn3 are not included in the frequency analysis of IS as present only in three cases (2 copies in #18, and one in #62; Table 1). For short, IS*200*/IS*605* is depicted in the figure as IS*200*.

analysis. Still, cumulative frequency distribution is suitable for finding out some reliable statistical trend on bird and fish ISs in mtDNA sequences (Figure 1, bottom). The statistical analysis of these and other possible differences will be presented below. Before that let us consider what kinds of ISs there are in the sample.

IS-Finder was able to detect by alignment a set of quite short ISs among the sampled 74 mtDNA nucleotide sequences. The ISs' lengths usually range from 17 up to 25 bp (in *L. obesus* mitogenome, 17-20 bp were met; Table 4, Column 4), although ISs themselves are known to be much longer (740-7,900 bp; Bacterial DB: http://www-is.biotoul.fr/is.html). Even in the example presented in this paper they vary in the limits from 1,372 up to 3,956 bp (Table 4, Column 3). Obviously, current data allow a conclusion on representation in 74 mitogenomes only sections of ISs, no natal ISs have been detected at all. The location of these IS sections changed widely, starting at CR and extending then nearly throughout the entire mitogenome. As exemplified for bullhead catfish, they were obtained at structural genes like *NADH-1, NADH-5*, *Cyt-b, Co-1* and at inter-gene spaces (Table 4, Column 2). The identity of aligned IS fragments to bacterial source IS is high, varying within 92-100% in the torrent catfish mtDNA (Table 4).

Repeatability of alignment technique in IS-Finder was checked in a different procedure. For this all 11 IS sections that have been obtained in torrent catfish (Table 4, Column 5) were aligned as target sequences with original ISs in a new IS-Finder search using the bacterial database. Such search showed that most of these IS sections aligned properly with natal ISs (Table 4, Column 7; Table 5). For 7 out of 11 ISs proper matches were obtained with the same IS type (Table 4, Column 7, underlined sequences; Table 5). However, 2 new similar matches were detected in other ISs and three IS*231K,* IS*Cth14* and IS*Cb2* had no back matches to itself (Table 4, Column 7). To map the location of IS sections in a natal bacterial sequence, the MS Word Find utility was applied (Find Motif utility of MEGA-4 gave same results). In this search 7 out of 11 target sections of IS sequences were located in originally defined ISs in bullhead catfish mitogenome as mentioned above (Table 5). Thus, it is possible to conclude that technique of IS search on target short sequences has 64-82% repeatability, i.e. 7 out of 11 or 9 of 11 IS match properly for data tested in Table 4 and Table 5. Judging on IS fragments detailed in Table 4 (column 5) and in Table 5, the ISs sections are look like in some attributes (TA or TAA presence) as inverted repeats in the analyzed mitogenomes. In other words, obtained IS elements are probably inherited as inverted repeats of real ISs that sometimes may be harbored by a mitogenome. As was noted above, not all ISs that were detected in original lists (Table 2, Table 4) match properly in repeated search. The reason for this may be stochastic. Some matches are find just by chance because many S-scores used in our data base were not so big and E-values not so small to avoid or minimize chance alone matches. Sample of real S- and E-scores are given in the Table 3. From these scores it is seen that most S-scores are within the limit 34–38, while many E-values are within 0.2–1.0. These E-values are much below E = 10 that accepted in IS-Finder by default when forming output list of ISs. If take data in Table 4 as representative, the number of error matches may reach up to 36%. If take probability estimates given in Table 2 as E-values related to S-scores, then chance alone IS occurrence in the database of the paper is 0.1. These estimations obviously showed that many (74-90% from both kinds of calculations) fragments of real ISs still existed in 74 birds and fish mitogenomes.

To address the second goal raised in the Introduction on the transmission routes of ISs through mitogenome, the ISs align scores for 60 variables (62 minus 2 grouping variables) were analyzed. These variables represent the IS families with the certain IS score as: (1) S-score and (2) E-value (all *i-th* scores with *i = 1..m*, where *m* is the IS ordinal number, were used to achieve the maximal power of statistical analysis). The scores were taken for each IS family selected in the database from output tables of IS-Finder, three of which are summarized as an example for two bird's, and one fish species (see Table 1). Selection of IS for the analysis was based on the availability of IS. So, ISs that most frequent in the database were used for calculations; those that unique or rare to be useful in the calculations were not selected. The entire set of the variables was of two types: type 1 and type 2, as represented by S-score and E- value. These so-called S- and E-variables are highly correlated: $r_p = 0.59 – 1.0$ ($n = 74$, $P < 0.01$). Thus, the final stage of the analysis included only scores of one variable type, the S-scores. There are three main outcomes of this analysis which summarized in Figure 2. Firstly, there is a statistically highly significant association between two groups of variables: grouping one (G1 and G2) and the rest variables, the set of which is represented by IS align S-scores, as defined in the Material and Methods and above (Canonical R = 0.8289, $X^2 = 104.48$, d.f. = 60, $P = 0.000034$). Secondly, all four groups comprised by two groups selected for birds and two groups selected for fish are quite well different in the scale of canonical variables (CV), as exemplified by CV scores of these groups which are marked by different labels (Figure 2). The classification precision of a certain IS to its own group of four defined is high and averaged to 85.14%. Thirdly, differentiation along the CV1 axis is stronger than along the CV2 axis as it follows from the roots removal; i.e. the statistics for the first root removed (CV1) comprised: R = 0.83, $R^2 = 0.68$, $X^2 = 104.48$, d.f. = 60, Lambda = 0.1574, $P = 0.00034$; for the second root removed (CV2): R = 0.71, $R^2 = 0.50$, $X^2 = 38.84$, d.f. = 29, Lambda = 0.5029, $P = 0.1049$. The CV1 root is mostly influenced by the following variables: IS*3*-S, IS*4*-S1, IS*21*-S, IS*256*-S, and IS*1380*-S1 with the correlation coefficients (CV weights)

Table 4. Location of aligned IS nucleotide sequences and their properties in the torrent catfish, *Liobagrus obesus* mitogenome (DQ321752) according to IS-Finder search at10.10.09

IS Sequence Name	IS Family/ Target Gene	Total IS Length, bp	Align-ment Length, bp	Location of IS Segment in the Query Sequence (complete mitogenome), bp limit: sequence order	Identity of Aligned mtDNA IS Section with Target Bacterial Sequence	List of IS Segments Find in BLAST IS-Finder Search Based on mtDNA IS Segments Given in Column 5**
IS*Sba20*	IS*21*/tRNA-Phe	2394	18	39-56: gttaagacgaaccctaga	18/18	IS*Sba20*
IS*Aar10*	IS*1380*/12S rRNA	1602	17	*312-328: ttcgtgccagccaccgc	17/17	IS*Aar10,* IS*Aar32,* IS*Aar11,* IS*Aar9,* IS*Mno1*
IS*Aar32*	IS*1380*/12S rRNA	1738	17	*312-328: ttcgtgccagccaccgc	17/17	IS*Aar32,* IS*Aar10,* IS*Aar11,* IS*Aar9*
IS*Tel2*	IS*200*/IS*605*	1675	18	1015-1032: acttggaataatcagggc	18/18	IS*Tel2,* IS*Fnu2,* IS*1628*
IS*231K*	IS*4*/NADH-1	3956	20	3012-3032: ttaaactatttattaaagaac	20/21	IS*Mst1*, IS*Fnu7*, IS*Cpe2*, IS*Cba1*
IS*Mba14*	IS*H3*/tRNA-Gln	1503	17	3911-3927: tattatggagataaaaa	17/17	IS*Mba14*
IS*Cth14*	IS*110*/NADH-2	1372	17	4686-4702: ctaatctcctcaacaaa	17/17	IS*Cth14*
IS*Cb2*	IS*4*/Co-1	1523	17	5644-5660: ttataattttctttata	17/17	IS*Cb2,* IS*Peth2*, IS*Mmy1*
IS*Mac7*	IS*200*/Co-1	1711	17	6333-6349: acagtaggaatagatgt	17/17	IS*Mac7,* IS*1165*, IS*Cwa1*, IS*Sto6*, IS*Cfe1*
IS*Bce1*	IS*1182*/NADH-5	1803	17	13370-13386: aaaaaccccaatcataa	17/17	IS*Bce1,* IS*Shes15*, IS*Rpa4*
IS*Mgi1*	IS*1380*/Cyt-b	1614	17	15389-15406: tcgggcaggtcgcctccg	18/18	IS*Mgi1,* IS*Arsp6*, IS*Blo8*, IS*427*

Note. *) These sequences have also the same location in the subject IS sequence: IS*Aar10* - 2ttcgtgccagccaccgc 282, IS*Aar32* - 266 ttcgtgccagccaccgc 282.
**) Only those ISs included in the list that have E-Score =< 1. Highlighted are sequences which location as IS sections mapped in bacterial ISs as shown in Table 5.

Table 5: Location of IS sections (bold and italic) of torrent catfish, *Liobagrus obesus* in ISs found in bacterial database by IS-Finder after repeated blast search against target IS sections

Sequence name: ISSba20.
Nucleotide sequence content:

TATTGCGCGACAATTAGCCTGACCGGTTCAGCGACAATTAGAATGGCCGGTTGATCTGATACATTCTGGACAAAGTTTCCAAGGGAGTTCTTTGTG
CCAGGTCGTAGAATTACAGATCAACAAATAAGGCTATTTATGTCTAAACGAAAAGATCATCCTCAAGTTACTGCTGCGGTTAAAGCGGGCATATCTG
AACGCTCTGCACGGCGTATTGAATCCGGCCAGCGACAATTAGGCCCTTCAAAACCTCGAAACTACCGCACTCGTACTGACCCACTAGAGCCTGTA
TGGGAACCTGTCGTCCTGCCGCTTCTACAACGGTCTGATACTATTACTCCTGTTGGGGTGTTTGACTATCTTATTGAAGAATATTCAGATGCTTTCC
CTGCAAA***GTTAAGACGAACCCTAGA***GCGACGTATACAAAAATGGCGACAGATAAATGGCCGAGACAAGGAGGTTATTTTCCGCCAAGTTAAACAA
TTGGGTCAGTTAGGAATTATGGATTTTACTTGGGCTGATTTCACCGTCACGATACAAGGAGTGGCTCTTAAACACAGATTATTTAACTATCGATTAC
CCGCAAGTGGTTGGAGCTACGCTGAAGTCGTATATGGTGGAGAGAGCTTTGTGGCCGTTGCTACAGGCTTACAAAATGCCTTCACACAATCAAAT
GGCGTTCCACAAGAAGTGAGAACAGATAGCTTGAGCGCCGCGTATAAAAACCATTCTAACGAAATATGGTTTACCGAACGATTTTCTGAATTATCAA
TGCATTATAGCTTCAAACCTTCAAAAAATAACACCGGCATTGCCCATGAAAATGGGGCGATTGAAAGTGCCAACAATCATCTGAAAAATCAAATACG
ACAAGCTTTGGCTATTCGTGATTCAAGTGACTTTGATTGCATAGACGAATATGAAGTATTCATTGATGAGGTCGTCCAAAGACGCAACCGCCGTATT
ATGCCACTTCTCATTGAGGAGCAACGACAATTACAACCTTTGCCTAAATTTGACAGTGCCAATTATGAAGTTCACCCAGTAAAAGTATCAAGCACCA
GTACTTTTCAGTTAAAACGAGTGACTTATTCAGTCCCATCTAGACTTGTTGGTGCAACATTACGCGTACATCTTTTCGATAAGACATTGGATATCTAT
TGCCAGGGCGTGCACACATCAACGCTCACCCGCGTGCATACATCGGCAAATCATCGAGGTCATCAAATTGATTACCGTCACTTAATCGGTGCGCT
GATGAAAAAACCACGAGCGTTTAGGGGGTGCCAATGGCGAGACCAACTCCTTCCAAATGAAGACTATCGTCAAATATGGAAAGGTATCGATGCCC
AATTAAGTGCTGACGAAGCGAGTCTTTATATGGTTAGGTTACTGAATATTGCCTGTAAATCAGAGCGAGAAGAGGCGGTAGGAAGATTTGTCCTTG
ATGGACTGAATAACGCTCAACTGCCAAGCATATTCGACTGTGAAGACCGTTTTTTAAAAGACGAAGAGTGGGAATACAACCCTCAAGTACAGCAAC
ATAGCTTAGCGTCTTATCAGCAGCTCTTCGACGGGGAGAATGGATATGTCAGTTGAAACATTACCCCTTCTTCTCAAAGAGCTTCGGCTAGCGAGC
TTACTTCCCCATTGGGAGGCGTTGGCAGAAAAAGCCCGAGAGCAACATTGGCCAATGGAGCGTTATTTAGCTGAATTATGCCAATTAGAACTGAGT
AGCAGAGAGCAAAAACGATTACAACGTGGTCTAAAAGAGGCGACGTTGCCAATAGGCAAATACCTTGATACCTATGATTTTAATGAAGTTGAAGGT
TTATCGAAGGAGCAGGTTTGGCATTTGGCCGAGCATGCGCAGTGGTTAAAGACGGGAGATAATATCTTGCTGTTCGGTGCTAGCGGTTTAGGTAA
AACCCATATTGCCGCAGGACTGGGTTATCGGCTTGTAGAGCAAGGGCATAGAGTCAAGTTTATGAGTGCGAGCTTACTTGTGCAGCAACTGCAAA
AAGCGAAAGAAGAGCTAAGGTTGTCGGAAGCCTTGGTAAAATTGGATAAATTCGCGGTTCTAATTTTGGATGATTTAGGCTATGTGCAAAAAAGCA
CAGAAGAAACGAGCGTATTGTTCGAGCTGATCGCGCATCGTTATGAAAGGCACAGTCTAATCATCACCTCAAATCAGTCATTCGAAGATTGGGATA
AGCTATTCAGTGATACGGTGATGACAGTAGCCGCAATCGATAGGTTGATCCACCACGCAAAGATCTTGCAATGCAAAGGAGAAAGTTACAGGCGA
AAAGAAGCACAAAACAAGCTAAATTAAACAGGACTTCAACCGGCCAAGTTAATTGTCGCTGGATCGGCCAATCTAATTGACGCGCTATA

Sequence name: IS*Aar10.*
Nucleotide sequence content

CCCGGGTATTTCATGAAGGCCGGTAGTCATTAACTGACTTAAAAGATGAAAAGGTGACCTGAACTGGGAAAATGGGAGTGTCTAAGCTTCCAAAAT
ACCCGTTCAAAGGATCACCTTTTCAATGAACCATTCTACCCACGTTTTCCCTGCCTTCTCGACCCAACTCACCGGCCAGGCCCTGGTCTCTCATGC
AGGGTTGTCGGTGCTGACCAGTTTCCTGAATGCCTTGGACTTCCGCAGCCTCTGCGAGAACCGGTTCAGCCAG***TTCGTGCCAGCCACCGC***AACC
CACCGTCCAGGCAAGATCCTCGGAGCACTGGCTCTATCGTTGGCGGCCGGCGGTGAACAAGCCACAGATATTGACCAGCTACGCACCGCACCAG
AACTCTTCGGCTCCGTAGCTTCTGATGCCACGGTCAGCCGATTCATGGGCCGGATCAAAGAACAGCCAGAAGCTTTCTCCTACGGGTTCGCCACC
ATGACCCGCAACCTGCGATCCAAAGTCTGGGCGGCAGCCGGAGCACGGAACCCCGCCCGGCTGGCCACGCCAGCCAACCCGCTGACCATCGAC
ATCGACGCCTCCCTGGTGCAGGTCCATTCCGAAAAAGAATCCAGCGCAGGAACGTATAAAGGCGGATACGGGTTCTCGCCGATGATCGCGATGG
CCGACTACGGCAAAGCCAATGGAACCGGCGAAGTCCTCGCGGTCCAGTTGCACCCGGGAAACCGGGGCGCGAATTCCGCCAAATCCCACATCG
ACGTACTCAACCAAGCGCTGGCGCAGCTGCCTGATGATTTCTACGACGAGCACGGAAACCTGTATCGAGAGAAGATCCTGGTCCGTACCGACAGT
GCTGGGTCCTCCCGGGAGTTCTTGCACTACCTGGATTCGTTGGGGATCCAATTCTCCACCTCGTACTCGCTACCGGTCATCAAGGAGCGGTTCAT
CCGGTGGATCGATGAGAAGAAATACTGGGAACCAGCGCTGACCGCTGACGGGCAGGAACGTGATGACGCGTGGGTGATCGACGCGAGCAAGGT
GATCGAGCTGAAGGACTACCCTCCAGGAACCCGGATCTATTTGCGGGCCGAGCCGTTGCATCCCGGCGCGAAAGCGACCTTGTTCGATACGGAC
GGGAATAGGGTGACTGCGTTCTTGACCAATAGCCCGCGGTTCAACGTGGCGTTCCTCGATGCCCGGCATCGTGCGCGTGGCCGGTGCGAAAACA
GGATCAAAACCCTGAAGAGCGCAGGGTTGGGCAAGCTGCCGTATTGGTCTTTTGCCGCGAACCAAGCATGGGCTGATCTGGCGATGTTCGCACT
GAATCTGGTGTCGTGGCTGCAGCTGGCCGTGCTACCCGGTGGTCATGACGCTTCGGTGTGGGATTTGAAGCGATGGCGGCACGGATTGCTGGC
CGCCTGAACTGGTGATTGTCCTGTTGTTTCTTTGGTTGCGAATGGTCATCACAATCCAGGCAAGTGGAGCCTGGCCCCCAGCGCCGGTGCGTCAA
GGCGCTATTTGGCATGGGCGGAATCAGTAACGGCCGCCGGGAATCCTGGGTATCGGAATTCGGCGGCCGTCAGAGTCCTCATGAAAAATTTGGG

Sequence name: IS*Aar32.*
Nucleotide sequence content

CCCGGGTATTTCATGAAGGCCGGTAGTCATTAACTGACTTAAAAGATGAAAAGGTGACCTGAACTGGGAAAATGGGAGTGTCTAATCTTCCAAAAT
ACCCGTTCAAAGGATCACCTTTTCAGTGAACCATTCTACCCACGTTTTCCCTGCTCTCTCGACCCAACTCACCGGCCAGTCCCTGGTCTCTCATGC
AGGGTTGTCGGTGCTGACCAGTTTCCTGAATGCCTTGGACTTCCGCAGCCTCTGCGAGGACCGGTTCAGCCAG***TTCGTGCCAGCCACCGC***TACC
CACCGTCCAGGCAAGATCCTCGGAGCACTGGCTCTATCGTTGGCGGCCGGCGGTGAACAAGCCACAGATATTGACCAGCTACGCACCGCACCAG
AACTCTTCGGCTCCGTAGCGTCGGATGCCACGGTCAGCCGATTCATGGGCCGGATCAAAGAACAGCCAGAAGCTTTCTCCTACGGGTTCGCCAC
CATGACCCGGCGCCTGCGCTCCAAAGTCTGGAATGCGGCCGGGCCGCGGAATCCCGCCCGGCTGGCTACGGCAGCCAATCCGCTGATTATCGA
CATAGATGCGTCGCTAGTGCATGTCCATTCCGAGAAAGAATCCAGCGCTGGGACGTATAAAGGCGGGTACGGTTTTTCACCGATGATCGCCATGG
CCGACTACGGCAAAGCCCACGGCACCGGCGAAGTCCTCGCCGTGCAATTGCGCCCGGGAAACCGGGGCGCGAACTCCGCCACCTCGCATATCG
AGGTGCTCGGCCAAGCGCTGGCTCAGTTGCCTGATGATTTCTATGACGAGCACGGGAACCTTCGTGGTGAGAAGATCCTGGCCCGTACCGACAG
TGCTGGGTCGTCCAGGGATTTTTTGCACCACCTGCATTCTCTGGGGCTCCAATTCTCCACGTCGTATTCGCTGCCGGTTCTCAAGGAGCGGTTCA
TCCGGTGGATCGATCAGAAGAAGTACTGGGAACCAGCCCTGGACGCTAACGGAGATCAGCGTGATGATGCATGGGTGATTGATGCAACGAACGT
CTTGGAGTTGAGGCAATACCCGCCTGGAACTCGGATTTACCTGCGTGCTGAGCCCTTGCATCCGGGGGCGAAAGCGAACTTGTTCGACACGGAC
GGGAACCGAGTCACCGCGTTCTTGACCAATGCGCCGCGGTTCAACGTCGCGTTCCTCGATGCCCGGCATCGGGCGCGGGGCAGGTGCGAAAAC
AGGATCAAGACGTTGAAGAATGCGGGGTTGGGCAAGCTGCCGTATTGGTCTTTCGCAGCGAACCAGGCGTGGGCGGACTTGGCGATGTTCGCG
GTGAATTTGGTGTCTTGGCTTCAACTTGCTGCGCTACCTGGCGGGCATGAGGCCGGCTGCTGGGATTTGAAGCGGTGGCGGTACCGGCTATTTT
CAATGGCCGGGAAAGTCGTCACCGGCGGCCGGCAACGCCGGCTGCTGATCGCCGAGAAGGCGCCTGAAGCGCAGCTGTTATGCCAGTTGCAGA
AGGGTATCGGTGAGTCATTCCAGCGGTGGCGGCACGGCGAGCTGGCCGCCTGAACTGGTGATTGTTCTGTTGTTTATTTGGTTTCGAAAGGCCAA
TTCAGTCCAGATAAGTGGAGCCTGGCACCCAGCGCTGGTGTTTCGGCGCTCTTTGGCGTGGGCGGAATCAGTAACGGCCGCCGGGAATCCTGG
GTATCGGATTTCGGTGGCCGTCAGAGACCTCATGAAAAATCTGGG

Sequence name: IS*Tel2.*

CAATGAAATGGGATACAAGCCCCGTCCTTTTAGGACGGCTTTTTTTGATTTCTCATGTATTCCTTAAGCACCTCTAACGGTGCTCCTCCAACAGATG
CCGCAAAATAACTAGGGCTCCAAAGTGATTCTTCATGGGGTTTTGGTAGTGCAGCTTTTCCATATCGACGACTAGATACGCCTTTAAGCGCATTTAC
GATCTGCGAAAGAGAAAGTTTAGGAGGGTACTCGATTAGCGCATGGACATGGTCTTCCTCGCCGTTAAATTCAAGAATCTGGAAATCCATCTTCTT
GGCGACTTCTCGAAACGATTTCTCGATCAGCTCTAATCCCTCAGCGCTAAGGACTGGCCGGCAATACTTAGTCACGCATACCAAATGAATCTTCAG
GTCCGTAACACTATGTCTCCCTTTACGAAGATGACTTGACATTGCCTGTAGACCAACCTATAATCTAGCTACAGACCAACTATATCACGACCAATGA
AAGCTAGATATAGGTATCGTTTCTATCCCACAGACCAACAACGACAACGCCTAGCTCAGTTGTTTGGCTGTGTCCGTGTGGTATGGAACGACGCAC
TGGCAATTTGTAAACAATCTGATTCATTGCCAAAAACTAGTGAATTGCAAAAGCTAGTAATTACCCAAGGGAAGAAAACACCAGAGCGGCAATGGTT
GTCTGATGTTTCTAATGTCCCGTTACAGCAGTCCGTTGCCGATTTAGGAGTTGCCTATAAAAACTTCTTCGATTCAATTAAAGGTCGGCGAAAAGGC
AAGAAGATCAACCCTCCTAGATTTAAAAAGAAGACAGAAAGACAATCCGCCAGATTTACTACCTATGGTTTTTCAATTAAAGGCGAAGAAGTTTATC
TAGCGAAGATTGGTAATGTCAGACCAATTTGGTCAAGAAAGCTTCCTTCTAAACCAAGCTCAGTTACGGTGATTAAAGATGCTGCAAATCGATATTT
TCTCAGTTTTGTCGTCGAGATCGACCCAGTTCATCAACCTGCAAGTAATCCTTCTATCGGCATTGATTTGGGAATTAAAGCGTTTGCTACTTTAAGT
ACGGGCGAAAAGATTGATGGCCCTGATTATTCCAAGTTAGACAAAAAGATCAGGCGAAAACAACGCAAACTGGCCCGTCAAGTTAAAGGAAGCCA
ACGGCGGGAGAAAACAAGACTGCAAATCGCTAAGCTTCATAGTCGGATTTCAGACATTCGCAGGGATTTCTTGCATAAAGTGTCTACCCGCGTGG
TTATCGAGAACCAAGTGATCACGCTGGAAGACTTGAATGTCTCTGGAATGGTGAAGAACCGCAAACTCGCTAGAGCAATTAGCTTGCAAGGTTGG
CGAGAATTCAGGGTTTTAGTTGAGGCTAAGTGCAACAAACTAAACCGTGAGTTTATTGTCATTGACCGATGGGAGCCAACCAGTCAGACTTGCTCA
AAATGCGGTTTTAAATGGGGCAAAGTCGATCTATCCGTTCGCTCGGTCGTATGTCTAAATTGTGGTACAGAGCACGATAGAGACGAAAATGCTGCA
AGAAACATAGAAAAAGTCGGGATAGGGCATTGCCACGACTATAAATGGACGCAGAGGGGGAGTAAGACTACTTCGGTAGCATCACCCAGTGAAGC
GTCAAGAATCATCGCACTTTAGGCCGGTGAGTATGTCAA

Sequence name: IS*231K.*
No specific match was found with target sequence in the motif finding.

Sequence name: IS*Mba14.*
Nucleotide sequence content

CAGTAGTTCCGAATAACTCAAGTTCCGAATAATGAATTTGGAACATTAGTAACTAATTATAATTACCCGAAAAAAAGTTTTATTAGATCATGGAATTCA
TTTTTAATGCCAACAAAAAACGATTCCATGACCTTAGATACAGCCAGCTTGCTGGCTGTATCTTCTGAATTGATTAGCAAATATAATATAATAACTTT
GCCAGAATCTGCAAATTATAAATGTCAGGACACTTTAAACATATTGTTGCATGCCGCAACATTCTCTACAAATTCTCTTGAATCTGCAAGTAATGATC
TGCAAAGAAAAAATCCTGACTTAAGAATACCTTCAGCAGATACTATTTTCAATTACATCAACGAAAATAAAATTGAAGATATTCTTTCCTCCTTTAGAA
AAATGAATCTTGAGCTTTTCAAGATGATGAAACTTGAAAACAAGATTCACGATATAGCTATTGATTTTCATGACATTTCC***TATTATGGAGATAAAAA***T
ACTCCTGGTATTAGGGGAATAAAACTAAAGAATGGTAGTTCATGGGGTAAATCGTTCTGTACTCTTGATATAATTGGAACTTCTCACTTAACACTTGA
CGTAATAGACATCAATAGTTTGAACAAGAATTATTCTCTTTTAATCGAATCTTTATTCAAAAGACTTCAAACGATAGGTGTAAAAACAGGAACAGTAT
ACCTTGATAGAGAATTTTTTAACACTGATGTGATTCCAAAATTGGATGAACTTAAAGTCAATTTTGTTATAGCAGTTAAATCGACTCAGGTTATCAATA
GAGAACTTAAAAATCATCAAAAAGAGTATGGAAATACTTCGACTATTTTTGAATATCAATTTCAGAAAGGAGGACCTTCTTTTAATGTTGTGGCAATA
TATAATGATGAAAAGAAAAAGTATCTTCTGTTTGCTACAAACAAAAAAGCAGAATCAATTGAGAAATTTGAAAAGATGATTCCTGAAGAATATAGAAA
ACGATGGAATATAGAAACTGGATATAGAGTGAAAAACGAGTTCAAGATACGATCATGTACAAAATCTCCAGTAGCGAGATTATTGTTCTTTATTATTC
AATGTATAATGTACAATGTGCTAAATATGCTAAAGTCAGTTCTAAAAATTACAGCTTATGAGCTAAAATCATTAATAAATGAAGATATAAACAAAGTTG
TAAGATATGGACTAAAGTCGCTGAATTTTATTCCTGTTAGTGTGTTACTGGATTGTTTGAGATGGGTCAATGAAGAAAGAAATCGAGTGCTACGCAC
TCGATTAACTATAATATAACTTATCTCAAATTAAAGTTTAAGAACCTAATAGCGTATGCTATAAAATTACTCAATGTGATCTGATTTTTTGAAAAATGTA
AGTGAAAGTTATAGCTTAAGAATTTTAAAATCTGAGCAATAATATAGTCTACTTTGAAAATTAATGGATAAAATGTAAGTTCAGTTTGTTTTTATCATTC
GGAAATCAGTCTGATCGGAACTACTG

Sequence name: IS*Cth14*
Nucleotide sequence content
No specific match was found with target sequence in the motif finding.

Sequence name: IS*Cb2*
Nucleotide sequence content:
No specific match was found with target sequence in the motif finding.

Sequence name: IS*Mac7*
Nucleotide sequence content
TTTATATAAAGAAGCGTAAAACCCCTGAGTCTTTAGCTCAGGGGATATAAGCGTCAACTTCAACCCTGATTCCGTATATAATATTCTGTGGCTTCTTT
ACTAACGTTTCCGATGGTAGAAGCAAAATATCCATCGCTCCACAATGTGTTTTCGCCCCTATAATACTTTTCCAGATATTCTTTTTGTGTTTTCCAAA
TTCTGTTCGTAGATTCATGTTTCTGTTTTCGGACAATAGATAGAACACTGACTTTCGTTTCGCTCTTAATTAAGAAATGGATGTGATCTTTGTCAGTTT
CCATTTCAAGAATTTCAAAGTTTGACCCTTTTGAAATATCAATCACAATTTGTTTGAGTTCTTCGCTGATTGGTTCAAGTATAGCTTTTCGATATTTGC
AAACAAAGATAATATGATACATTAACAGAAATTTGCTATGATTCCTTGTCTCATATTTCATATTTACAAAAAAATATATGAATTGTTAAAGCATATGATC
TCTTTGTATGATGAAAGCGTTCAAATTTAGGCTCTATCCTACAACTACACAAGCTGATCAACTCAATCAACATATAGGTAGCTGTAGGTTTGTCTATA
ATTGGGCACTTGATCAGAAAATTAAAACATACGAACAAACAGGGAAATCAATTTCCAGATTTGACTTAAACAAATTAATTCCAGCTTTGAAGACCTCT
AACGAATGGTTAGGAGAAGTCAATTCTCAATCATTACAGGGGATGACTAAGCAGGTTGAGTCTGCTTTCACTCGATTTTTCCGAGAGAAGAATGGA
TTTCCGAAGTTTAAGTCAAAGAAAAATCCTATCCAGTCTTTTCCCGTACCCCAACACTATTCTGTGAATTTTGAAAACCATACCGTTAAACTTCCTAA
AATTGGTGAGATTAAAGCAGTTCTTCACCGAAAATTTGATGGAGAGCTTAAAACAGCTAGGGTATCAAGGACTTGTAAAGGGCATTACTACATCAG
CGTCCTTGTTGAAGACGGTAAAGAACTTCCTGAAAAACAGGGATTCTCAGAATCAACA***ACAGTAGGAATAGATGT***GGGTATCAAAGATTTTGCCGT
CCTTTCCACAGGAGAAAAGATTGAGAATCCTAAATACCTGAAAAACTCTTTGCAAAGGTTAAAGGTCCTTCAGAAAAGAGTATCAAGGAAACAAAAA
GGCTCTAAAAATAGGGCAAAAGCCAGACAGAGGCTTGCTGTACTCCACGACAAAATAACAAATCAGAGGAATGACTTCCAGAACAAACTCTCTTTT
AGACTTGTTAGCGAAAACCAAGCTGTAGCTCTGGAAACTCTGAATGTTAAAGGTATGGTTAAAAATCATCATTTGGCACAGGCTATAAGTGATTCTG
CATGGAGCAGCTTTGTAACAAAGTTAGAATATAAAGCAGAATGGTTTGGAAAAGGCGTATTAAGGATAGGACAATTTGAACCTTCTTCTAAGCTCTG
TAACGTGTGTGGATACCATAATAAAGAGCTTCAGTTAAAAGACAGAGAATGGACTTGTCCTGATTGTAAAACAAAACATGATAGAGACATAAATGCC
GCTATCAATATCAAGAAATTCGCTCTCATAGATCAAAATCTAATTGGATTGTAACACCTACGGAACGTGGGGAAGAGCTTGGGGACTTGCCCTCAA
TAGAGGGAAGAATGAACCAAGAAGCAACTCAGTCTTTAGCTGAGTGGTAGTTCAC

Sequence name: IS*Bce1*
Nucleotide sequence content
GAGGTTGTCCCAAAAGTAGCTGAATAGCTACTTTTGGGCAACCTCTTTCTTTACGGAAAATATAAAAAAGAATCGTCCATTTTTGATATAATTAAGTC
ACCACAACCAAATCATCAAAAGAAAGGAACGATTCTTTTGTGTTTATAGATTATAACAAAAACCAACTGACTTTTCCACTAGATGTTGAGGTTCATAT
TCCTGAACATCATCTTTCACGTGTGGTTCATACGGCTGTAGAAGCGATGGATCTCTCTATCCTCTTATCACGTTATCCAGGTGGTGGCAGACCACC
TTATCACCCAAAAATGATGTTGAAAATTTTATTATATGCCTACGCAAAAGGAATCTATTCTTCTCGAAAAATTGAAGAACAACTCCATGAAAATATCCA
TTTCATGTGGCTTTCTGATGAAAAAACACCAGACTTTCGTACGATCAATCGTTTTCGATCGGAGCGCATGAAAGACGTGATTTATGAAACTTTCTTC
TCTCTCATCGACCTGTTGAAACAAGAAGGGCTTGTCAAACTCGAAGATTATTTTCTAGATGGAACAAAAATCGAGGCAAATGCCAATCGATATACGT
TTGTATGGAAAAAATCCACAGAAAAATATGATCAAAAACTAGATGAGAAGTACCAAAAAATTGTCGCGACAATTGAACAAGCAAGTGAAGACGATCA
GAAAGCGGAATTAGAAGAAGAACAAATGAAAGATACCTGTTCGGTCACTTCAGAAAAAATTAAACAAACCATTCAAACCTTAGAAAAACGCCTACAA
GATGAACCGAAAAATAAAACGATGAAAAAAGCAAAGCGTCAGTTAGAAAATGATCTTCTTCCGCGCAAAATAAAATACGAAGCTCATAAAGAAACAT
TTCAAGAAAGAAATAGCTTTTCGAAAACAGATACAGATGCGACCTTTATGCGGATGAAAGAAGATCACATGAAAAATGGTCAACTAAAACCAGGGTA
CAATATACAAATCGGAACCGAGGATCAATTTATTACAGGGTTTAGTATTCATCAAAAAGCGGGAGATACAACCTGTTTGATTCCTCATTTTGAAATAT
TTGAGAAGTACAATCGTGTAAAGCCTCAAAATGTAGTTGCCGATTCTGGTTATGGAAGTGAAGAAAACTATGCATTTTGTGAAGAGAAAAAAATGAA
AGCATATGTGAAATACAATACGTTTGAAAAAGAACAAAAGAAGGCGTGGACGAAACAAATAGGTCGCGTGGAAAATATGTTTTATGAGGAAGAACT
AGATGAATGGATTTGTGCAAATAAAAAACGCCTTATTTTTTCACATGAAAGCAAACAACAAACAAGCAATGGTTATACTTCTATCAAAAGAACCTATC
GTTGTACAGAATGTTTAGGCTGTCCATTTCAAGAAACATGTGCCAAAGGAAAAGAAACAAAATCTATCAGTGTCTCTGTCAAAAATCAAAAACAGCG
AAAAGAAGTACGCGAACGTTTACATACAAAAGAAGGTAAAGAATTATATGGAAAGCGAAAAATAGAGGTGGAACCCGTATTTGGTCATATTAAATAC
AATCGTTTCTTCCAACGTTTTTCCTTACGAGGCCTTTCCAAAAATCGACTGGAATGGGGTCTTATTTGTGTTGCCCATAATTTACGGAAATGGACAA
CAACAGCTCAATCAAAAGCGA***AAAAACCCCAATCATAA***GTAGAAAAACAATGGGTTTTCTATTTATGATTGGGGTTTTTATCCATATAAAGGTGAAA
TATAAGAAATTGTCTCCAAAAGGTCGTTTTTAACGACCTTTTGGGACAACCTC

Sequence name: IS*Mgi1*
Nucleotide sequence content
CCTCGACGTTTCATGAATTGGTCGGTGACGCGGGCCTGTAATGCACGGAAGTGCCTGTCCTAGTTGAGAACATTGACTTGCTAAGGGTCAAGGTT
CGCCAAGAGGAAAGGCACTCCGTAGATGAAGCGTAGCGCGGTGTGTTCGGTAGTGGTGGATTCGGGTCGCGAGTCGCTGGTGTCGTCGGCCGG
CGGGCTGTTGTTGGCCCGGACTCTGCACTGCTCCGGACTTGAAAAGGCCATGTCAGCGGCATTAGCGCCGTGGCGAGCACCGCGTGCGGTCCA
CGACCCGGCCAAGGTCTTGACTGATATGGCGATCGCAGTCGCTCTCGGCGGTGACTGCGCCGCTGATGTCGCGGTGGTCCGGGCCCAGCCCGA
GCTGT***TCGGGCAGGTCGCCTCCG***ATGCCACCGTGTCCCGACTGATCGCCGCACTCGCCGACGATGTCGATGCTGCTCTCACGGCGATCCGCAC
CGCGCGGGCACTTGCTCGTCAGCGTGTCTGGCGTCGTCAACGACCGTTGCCCGGCGCGGGCAACCAGGTGATCATCGACCTGGATGCCACCCT
GGTGACCGCCCACAGCGACAAAGAAGGTGCAACACCAACCTTCAAGTACGGGTACGGGTTTCACCCCATGCTGGCGTTCGTCGACCACGGCAGC
CACGGATCCGGCGAAGCGTTGGCAGGCCTGCTGCGACCCGGCTCAGCCGGGTCCAACAGTGCAGCCGACCACATCAGCGTCCTGGACGCCGC
CCTGGCGCAGCTTCCCGACCATGAGCGTGCCCAGGTCCTGGTACGGACTGACACCGGCGGTGGGGTCAAAGACTTCCTGCACCACATCACCGA
CCTGAAACTGCAGTACAGCGTCGGGTTCTACGGCATGCCCCCGATCGTCGAAGCCCTCAATCGGGTGCCGCCCCAGGCGTGGCGGGCTGCCCT
CGACGGCGATGGCGCACCGCGCGAGGGTGCCCAGGTCGCCGAGCTGACCCGCTACCTGCCCGACACACTGCGCGGCTGGCCGGCAGGGATG
CGGGTCATCGCTCGCCGTGAACGGCCCCATCCCGGCGCCCAGCTGCGCCTGACCGACGACAACGGGTGGCGGATCACCTGTTTTGCGACCAAC
ACCCCAGGCTGGTCGATCGCCGATCTGGAAGTTCGGCACCGTCAACGTGCTCGTGCCGAGGACCGTATCCGCAACCTCAAAGACACCGGCCTGA
CCAACCTGCCGTTTCACGGCTTCGACCAGAATCAGCTCTGGCTCGAAATCACCCTGCTGGCCGCTGACCTGCTGGTCTGGACCCAAGTCCTGGC
GTTCACCGCACACCCGGCCCGACGTTGGGAACCCAAACGGCTACGCCTACGACTGCTGGCCGTGGCCGGACGGATCATCCACACAGGACGGCG
TCGACTCCTGCGTCTGCCCCGCGGCTGGCCATGGAGCGACCTCATCGAAACCGGCTGGACTGCACTGCAAACCGCCTAGCAACACTCACCAAAC
CGACGACCAAGGACCCAGGAGCACCGGCGACACGCAACGCCGGAACTCCACGCCGCCCGTCATCGCAACCGACTTGCCACACCCAGAACACAC
ATCAACCACCGGCCACACGAAAGATCGAGG

Note. ISs in which match with target sequence was found are highlighted in Table 4 with gray shadow as well.

0.31, 0.38, -0.26, 0.30, 0.39, 0.45, respectively. The CV2 root is mostly influenced by the following variables: IS*3*-S1, IS*30*-S, IS*5*-S, IS*5*-S1, IS*481*-S, IS*91*-S, and IS*1595*-S. Visually, as designed with bold line types, greater dissimilarity of 2 clusters is seen along the CV1 axis, and it is represented by individual IS scores of two groups, birds and fish (Figure 2).

To support the results above, the analysis of different variables, i.e., the frequencies of 25 IS families (Figure 1, bottom) was done. For this analysis, the data on the IS content (Table 2, Column 5) were turned into frequencies. As with IS align S-scores, the blank cells in frequency table were treated as zero (frequency = 0); given IS got score 1 if met once, then score 2 if met twice, etc. Summary of such frequency distribution is given in Figure 1 (bottom). Frequency distributions are also not normal for all IS members ($P < 0.05$) and cumulative frequencies for bird and fish species are well different (Figure 1, bottom). Student's t-test obtained frequency differences between these two groups for IS*30*, IS*256*, IS*630*, and IS*1595* ($0.001 < P < 0.01$). Multi-dimensional Hotteling's criterion (T^2) showed that combined differences in IS frequencies between birds and fishes are statistically significant: $T^2 = 78.12$, F = 2.08, d.f. = 25; 48, $P < 0.014$. Because deviation from normality of frequency distributions it is preferable to use another and more robust statistics. For this, the canonical analysis with IS frequencies was performed first. This analysis showed that correlation between two groups of variables tested (first is G1, G2 and second is 25 IS members) is highly statistically significant: Canonical R = 0.77, $X^2 = 76.79$, d.f. = 50, $P = 0.0088$. At next step, ANOVA (main factors mode) was applied to CV roots to confirm above results. The main effect on G1 variable (or factor 1) indicating bird vs. fish difference in IS frequencies is as follows: Wilks $\lambda = 0.42$, F = 49.13, d.f. = 2; 70, $P < 0.0001$. Also significant statistically is the main effect on G2 variable (factor 2) indicating food chain difference in IS frequencies: Wilks $\lambda = 0.57$, F = 26.69, d.f. = 2; 70, $P < 0.0001$. More relevant MANOVA (interaction mode) showed that certain groups are also different in frequencies within 2 main factors: Wilks $\lambda = 0.85$, F = 5. 96, d.f. = 2; 69, $P = 0.004$.

All provided statistical effects can be interpreted as a source for two main conclusions. 1. Since birds and fish separately form more compact clusters along CV1 (the bird cluster fall in the negative CV1 scores and the fish fall in positive CV1 scores; Figure 2) and differed in IS frequencies, it can be concluded that there exist a maternally-dependent pathway for transmission of IS sections detected among the analyzed species groups and it seems to be stronger then an alternative one. 2. An alternative or horizontal pathway for ISs transmission among bird and fish species also has been revealed (ISs frequencies in food chain are different, $P < 0.0001$ or $P = 0.004$; see above), although it is statistically weakly supported in the first set of variables that represented by original S-scores ($P \sim 0.1$ and the factorial impact is nearly 50%; see statistics above).

To the best author's knowledge, the ISs presence in mitogenome, even as their sections or fragments, has not yet been detected. However, big sections of mtDNA were described as insertions in human nuclear DNA (Thomas et al., 1996; Willet-Brozick et al., 2001). Also a transportation of mtDNA CR into nuclear genome of several duck species was reported (Sorensen, Fleischer, 1996). Thus, potentially reciprocal way of transmission, i.e. from nuclear DNA elements to mtDNA might also be open. Nowadays, in bacterial genomes ISs are well known (Mahillon and Chandler 1998; Craig et al., 2002; Siguier et al., 2006b; Wagner et al., 2007; Wagner and Chaux 2008). Despite a relatively small size of Prokaryotic DNA, there is enough space there to harbour extra DNA elements, like ISs. No surprise having a variety of vectors for between-genome transportation,

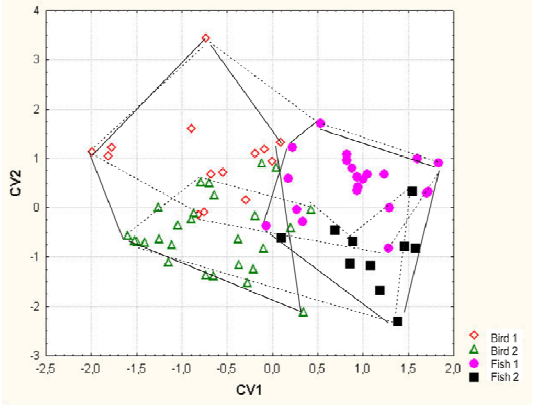

Figure 2: Bivariate plot of canonical variable (CV) roots obtained from IS variables as the align scores for 74 complete mitogenomes of 43 bird and 31 fish species. CV roots distribution are significant statistically: P = 0.00003 (see text for details). CV1 root is mostly influenced by the following variables: IS*3*-S, IS*4*-S1, IS*21*-S, IS*256*-S, IS*1380*-S1 with the correlation coefficients (CV weights) 0.31, 0.38, -0.26, 0.30, 0.39, 0.45, respectively. CV2 root is mostly influenced by the following variables: IS*3*-S1, IS*30*-S, IS*5*-S, IS*5*-S1, IS*481*-S, IS*91*-S, IS*1595*-S. Solid lines visualize gender-related IS transmission (from birds to birds, from fish to fish), broken lines show possible horizontal IS transfer throughout food chains.

including bacteriophages and plasmids, that mitogenome holds some ISs or their sections; up to 15 as was detected in our database (see Table 1, Table 4-5). As have been discussed in previous paragraphs, at least half of them should be real ISs' fragments. It is also evident that no full length ISs present in bullhead catfish and likely others mitogenomes. The reason for late seems is too short nucleotide space available in complete mitogenome and thus its limited flexibility. Most sequences that were analyzed in the paper are coding sequences. Thus, if mtDNA sequence harbour IS, even as its fragment, then this may broke ORFs and decrease fitness of a particular gene product and entire ribosome. However, fitness decrease of a ribosome itself may be negligible considering wide-spread heteroplasmy and so existence a source of normal gene sequence copies and their functional products in the same ribosome.

In our database as well as in bacterial genomes (Wagner et al., 2007; Wagner and Chaux, 2008) different ISs show a skewed and patchy distribution type, where most genomes carry no members of given IS family, and a very small number of genomes carry many members. Our data shows that only catfish mitogenome holds several copies of one IS: IS*1380* family and eight IS*Mgi1* copies. Anyway, the presence of IS fragments among 74 mitogenomes that were subjected to the analysis in the paper seems doubtless. Also, the difference in S-score values and frequencies of ISs looks realistic despite weak analytical properties of such database. Especially quite reasonable are differences between bird and fish part of IS members that were supported both by IS S-scores and IS frequencies. Such data may be interpreted as predominately maternally-dependent IS transmission. Alternative, horizontal or lateral transmission is also possible according to data presented in the paper. This topic requires special, more wide-scale consideration.

As discussed above it is probable that horizontal transfer of IS in nature would not be a surprise in a view of a number and a variety of autonomous extra chromosomal elements such as bacteriophages and plasmids, which can serve as vectors. One of the known evidence is, for instance, isolation of identical IS*6* family members from *Mycobacterium fortuitum* and *Flavobacterium* (Arthrobacter) *sp*. IS*6100* (Kato et al., 1994). These and newest observations clearly supports the idea that horizontal transfer occurs among Prokaryotes in nature (Wagner 2009), although it may be quite rare for distant lineages (Wagner et al., 2007; Wagner and Chaux, 2008). An analysis of the nucleotide sequences of IS*1*, IS*3*, and IS*30* from the ECOR collection and from other related enteric bacteria shows that each type of IS is highly conserved within *Escherichia coli* (Lawrence et al., 1992). Since the degree of sequence divergence of several chromosomal genes within these clonal lineages was significantly larger, it was concluded that the IS had a high turnover and rapid movement. Moreover, strains carrying one type of IS element also tend to carry other types. This observation is consistent with our data and also with the idea that multiple ISs can be delivered by a single vector, for example a transmissible plasmid or phage (Sawyer et al., 1987). This finding is corresponds as well with our observation on a predominated transmission impact in more close relatives: birds, group 1 vs. group 2, fish, group 1 vs. group 2 and weaker in the alternative direction along food chain pathway (Figure 2, and text above). Numerous IS horizontal transmissions were detected in Eukaryotes in plant and animal species

(Pritham 2009) and in Prokaryotes as discussed above. However, to the author's best knowledge, there have been no records yet on ISs presence and their horizontal transmission within mitogenome. The horizontal transfer was judged as rare event in historical time for distant lineages even in Prokaryotes at any rate (Wagner et al., 2007; Wagner and Chaux 2008). Thus, for our case with birds and fish, the diet preferences, as defined with G2 variable, may be only potentially the source of ecologically dependent IS transmission throughout food chains. Certainly, we need a bigger data set and more powerful statistic tools to analyse this phenomenon and to draw more definite conclusion on this point. BLAST search performed by IS-Finder itself has internal difficulties dependent on the databases available at a time and this is another source of the uncertainties in the analysis. Still, even if horizontal transfer occurs rarely and that the latter fact itself and the mechanism of transmission are obscure now in mitogenomes analyzed, such potential horizontal transmission in one food chain (fish → bird) inevitably raise question on a potentially harmful impact in other organisms' chains, like fish → human for instance. The latter is quite possible in many countries where crude marine food (sushi and sashimi) is normal in an every day diet. Thus, current paper may be important even as an attempt to attract researches' attention to this phenomenon.

ACKNOWLEDGEMENTS

We are very thankful for proofreading of the manuscript and useful comments to Mrs. I.A. Barsegova. This work was supported in part by Far Eastern Branch of Russian Academy of Science grants: # 09-I-P23-07.

REFERENCES

Avise JC (2000). Phylogeography: The History and Formation of Species, Harvard Univ. Press. Cambridge.

Craig N, Craigie R, Gellert M, Lambovitch A (Eds) (2002). Mobile DNA II. ASM Press. Washington DC.

Galas DJ, Chandler M (1989). Bacterial insertion sequences. In: Berg D.E. and Howe M.M. (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C. p. 109–162.

Kartavtsev YP, Lee Y-M, Jung SO, Byeon HK, Son Y, Lee JS (2007a). The complete mitochondrial genome of the bullhead torrent catfish, *Liobagrus obesus* (Siluriformes, Amblycipididae): Genome description and phylogenetic considerations inferred from the *Cyt b* gene. Gene 396: 13-27.

Kato K, Ohtsuki K, Mitsuda H, Yomo T, Negoro S, Urabe I (1994). Insertion sequence IS6100 on plasmid pOAD2, which degrades nylon oligomers. J. Bacteriol.. 176: 1197-1200.

Kim IC, Jung SO, Lee YM, Lee CJ, Park JK, Lee JS (2005) The complete mitochondrial genome of the ray fish *Raja porosa* (Chondrichthyes, Rajidae). DNA Seq. 16: 187-194.

Kogelnik AM, Lott MT, Brown MD, Navathe SB, Wallace DC (2005). MITOMAP: a human mitochondrial genome database – 1998 update. Nucleic Acids Res., 26: 112-115.

Lawrence JG, Ochman H, Hartl DL (1992). The evolution of insertion sequences within enteric bacteria. Genetics 131: 9-20.

Mahillon J, Chandler M (1998). Insertion sequences. Microbiol. Mol. Biol. Rev. 62: 725–774.

Nagase M, Aimi T, Suginaka K, Kitamoto Y, Morinaga T (2005). Complete mitochondrial DNA sequence of the Japanese flying fish *Cypselurus hiraii*. Fisheries Sci. 71: 914-923.

Nei M (1987). Molecular Evolutionary Genetics. Columbia Univ. Press. New York.

Nohara M, Nishida M, Miya M, Nishikawa T (2005). Evolution of the mitochondrial genome in cephalochordata as inferred from complete nucleotide sequences from two epigonichthys species. J. Mol. Evol., 60: 526-537.

Pritham EJ (2009). Transposable Elements and Factors Influencing their Success in Eukaryotes. Journal of Heredity 100(5): 648-655.

Radice AD, B. Bugaj, Fitch DH, Emmons SW (1994). Widespread occurrence of the Tc1 transposon family: Tc1-like transposons from teleost fish. Mol. Gen. Genet. 244: 606–612.

Robertson HM, Lampe DJ (1995). Recent horizontal transfer of a mariner transposable element among and between Diptera and Neuroptera. Mol. Biol. Evol., 12: 850–862.

Sawyer SA, Dykhuizen DE, DuBose RF, Green L, Mutangadura-Mhlanga T, Wolczyk DF, Hartl DL (1987). Distribution and abundance of insertion sequences among natural isolates of Escherichia coli. Genetics 115: 51–63.

Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M (2006a). ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. (Database Issue) 34: D34–D36.

Siguier P, Filee J, Chandler M (2006b). Insertion sequences in prokaryotic genomes. Current Opinions in microbiology 9: 526 – 531.

Sorensen MD, Fleischer RC (1996). Multiple independent transpositions of mitochondrial DNA control region sequences to the nucleus. Proc Nat. Acad. Sci. USA 93: `15239-15243.

STATISTICA 6.0 (2001) Statistica for Windows: Users Guide, StatSoft, Tulsa.

Thomas R, Zischler H, Paabo S, Stoneking M (1996). Novel mitochondrial DNA insertion polymorphism and its usefulness for human population studies. Human Biology 68(6): 847-854.

Wagner A, Lewis C, Bichsel M (2007). A survey of bacterial insertion sequences using IScan. Nucleic Acids Research 1–10. doi:10.1093/nar/gkm597

Wagner A (2009). Transposable elements as genomic diseases. Mol. BioSyst. 5: 32–35.

Wagner A, Chaux N (2008). Distant horizontal gene transfer is rare for multiple families of procariotic insertion sequences. Molecular Genetics and Genomics 280: 397 – 408.

Willlet-Brozick JE, Savul SA, Richey LE, Baysal BA (2001). Germ line insertion of mtDNA at the breakpint junction of a reciprocal constitutional translocation. Human Genetics 109: 216-223.