




Article

# Identification of Mitochondrial DNA (NUMTs) in the Nuclear Genome of *Daphnia magna*

Krzysztof Kowal <sup>1</sup> , Angelika Tkaczyk <sup>1</sup>, Mariusz Pierzchała <sup>2</sup>, Adam Bownik <sup>3</sup>  
and Brygida Ślaska <sup>1,\*</sup>

<sup>1</sup> Faculty of Animal Sciences and Bioeconomy, Institute of Biological Bases of Animal Production, University of Life Sciences in Lublin, Akademicka 13 Str., 20-950 Lublin, Poland; krzysztof.kowal@up.lublin.pl (K.K.); angelika.tkaczyk@up.lublin.pl (A.T.)

<sup>2</sup> Department of Genomics and Biodiversity, Institute of Genetics and Animal Biotechnology of the Polish Academy of Sciences, Postępu 36a Str., 05-552 Jastrzębiec, Poland; m.pierzchala@ighz.pl

<sup>3</sup> Faculty of Environmental Biology, Department of Hydrobiology and Protection of Ecosystems, University of Life Sciences in Lublin, Dobrzańskiego 37 Str., 20-262 Lublin, Poland; adam.bownik@up.lublin.pl

\* Correspondence: brygida.slaska@up.lublin.pl

Received: 12 October 2020; Accepted: 16 November 2020; Published: 18 November 2020



**Abstract:** This is the first study in which the *Daphnia magna* (*D. magna*) nuclear genome (nDNA) obtained from the GenBank database was analyzed for pseudogene sequences of mitochondrial origin. To date, there is no information about pseudogenes localized in *D. magna* genome. This study aimed to identify NUMTs, their length, homology, and location for potential use in evolutionary studies and to check whether their occurrence causes co-amplification during mitochondrial genome (mtDNA) analyses. Bioinformatic analysis showed 1909 fragments of the mtDNA of *D. magna*, of which 1630 were located in ten linkage groups (LG) of the nDNA. The best-matched NUMTs covering >90% of the gene sequence have been identified for two mt-tRNA genes, and they may be functional nuclear RNA molecules. Isolating the total DNA in mtDNA studies, co-amplification of nDNA fragments is unlikely in the case of amplification of the whole tRNA genes as well as fragments of other genes. It was observed that *TRNA-MET* fragments had the highest level of sequence homology, thus they could be evolutionarily the youngest. The lowest homology was found in the D-loop-derived pseudogene. It may probably be the oldest NUMT incorporated into the nDNA; however, further analysis is necessary.

**Keywords:** mtDNA; NUMT; water flea; BLAST

## 1. Introduction

*Daphnia*, commonly known as the water flea, are small crustaceans usually inhabiting freshwater ponds and lakes on all continents of the globe. It has long been used as a model for the elucidation of animal responses and adaptations to environmental changes [1]. It has been also used in diverse biological research areas such as ecology, ecotoxicology, evolution, and reproductive biology due to its important position in the aquatic food chain, a high degree of phenotypic plasticity, and cyclical parthenogenesis responding to environmental stimuli [1–7]. Its sensitive behavioral and physiological responses are parameters used as biomarkers of the effect induced by various substances [1–8]. Thus, it has been used for reproduction tests, acute toxicity studies, and chronic toxicity tests in the OECD Guidelines [9,10]. The current state of knowledge of the nuclear genome of *Daphnia magna* is based on reports by Routtu et al. [11,12], Dukić et al. [13], and Lee et al. [14]. Low- and high-density

genetic linkage maps were obtained, in which they assembled the whole genome sequence of *D. magna*. Specific genetic markers from a high-resolution genetic linkage map of *D. magna* Xinb3 were evaluated in toxicological studies by Korea Institute of Toxicology (KIT) [14].

Genomic resources are steadily being developed for many species of the genus *Daphnia*. In particular, a database of around 12,000 expressed sequence tags (EST) is currently available (<http://wfleabase.org>) [15], providing a useful resource to isolate polymorphic genetic markers in this species. However, there is no information about pseudogenic sequences of mitochondrial origin (NUMTs) in the *D. magna* genome.

Nuclear DNA sequences that are homologous to the mitochondrial genome are often referred to as mitochondrial pseudogenes, or NUMTs [16]. NUMTs may differ in length and be as large as the full length of the mitochondrial genome [17]. It has been reported that the NUMT length is positively correlated with the genome size, suggesting potential roles of non-coding DNA gain and loss in NUMT accumulation [18]. NUMTs have been documented in almost all eukaryotic genomes studied [19]. The transfer of mitochondrial DNA sequences into the nuclear genome is an ongoing evolutionary process [20], which has markedly influenced the evolution and function of eukaryotic genomes [19]. Thus, NUMTs are good materials for studying the evolution of nuclear sequences without selective constraints [21]. However, because of their homology, NUMTs may confound mtDNA studies, as the NUMT co-amplification product could interfere with sequence analysis [22].

This is the first study in which the *D. magna* nuclear genome deposited in the GenBank database was analyzed for pseudogene sequences of mitochondrial origin. The present study aimed to identify NUMTs, their length, homology, and location for potential use in evolutionary studies and to check whether their occurrence causes co-amplification during mitochondrial genome analyses.

## 2. Results

Bioinformatic analysis showed 1909 fragments of the mitochondrial genome, of which 1630 fragments were located in ten linkage groups of the nuclear genome of *D. magna*. The other fragments were localized in scaffolds which were used during genome sequencing. All the NUMTs found in this research are listed in supplementary materials (Table S1). The total length of the NUMT sequences in the linkage groups corresponded to the number of fragments on individual LG (Table 1). The total length of NUMTs in the *D. magna* genome was 44,391 base pairs (bp), which accounted for 0.042% of the length of the nuclear genome. Their percentage content was 0.037% in the longest linkage group 2, in which 228 NUMTs were identified, and 0.047% in the shortest linkage group 10 (Table 1). The most frequently occurring fragments of the mtDNA sequence in the nuclear genome included *ND2* (115), *ND3* (113), *TRNA-CYS* (110), and 16S rRNA (105). However, the highest number of NUMTs was observed for the non-coding area, i.e., the D-loop (147) (Table 2). The lowest numbers of mtDNA fragments found in nDNA were observed mainly for genes encoding tRNA molecules: *TRNA-PRO* and *TRNA-MET* (6), *TRNA-TYR* and *TRNA-ASN* (4), and *TRNA-SER1* (2). The highest numbers of mtDNA fragments were recorded on LG2-228, and the lowest—on LG8-134 (Table 2).

The longest fragments of the mitochondrial genome present in the nuclear genome were observed for the D-loop (182 bp), *ND4* (108 bp), *ND3* (99 bp), and *ND5* and *COX3* (94 bp each) (Table 3). The 182-bp fragment of the D-loop constituted 63% of the entire sequence of this region. In contrast, in the case of the other protein-encoding genes, the fragment size ranged from 4% (*CYTB*) to 32% (*ATP8*). In turn, NUMTs were recorded among genes encoding tRNA, constituting over 90% of the mtDNA gene sequence: *TRNA-ARG* and *TRNA-THR* (95% each) and *TRNA-GLU* (97%). All the analyzed sequence fragments had a minimum length in the range of 16–25 bp (Table 3).

**Table 1.** Percentage content of NUMTs in the linkage groups of the nuclear genome.

Linkage Group (LG)	Physical Length (bp) from Lee et al., 2019	Sum of NUMTs Lengths (bp)	Percentage Content of NUMTs on LG
LG1	14,058,888	5.063	0.036%
LG2	16,351,056	6.071	0.037%
LG3	11,081,246	4.389	0.040%
LG4	10,002,879	4.147	0.041%
LG5	10,116,075	4.600	0.045%
LG6	9,588,688	3.997	0.042%
LG7	10,149,764	4.829	0.048%
LG8	9,006,911	3.533	0.039%
LG9	8,299,553	3.966	0.048%
LG10	8,061,327	3.796	0.047%
Total	106,716,387	44.391	0.042%

**Table 2.** Distribution of mtDNA gene fragments in the linkage groups and the sum of fragment length.

mtDNA Sequence	LG										Number of NUMTs (100% Identical) *	Sum of Gene Fragments (bp)
	1	2	3	4	5	6	7	8	9	10		
<i>TRNA-GLN</i>	3	6	5	7	2	5	3	2	5	4	42 (15)	966
<i>TRNA-MET</i>	2	2	1	1	-	-	-	-	-	-	6 (6)	102
<i>ND2</i>	12	18	15	12	12	9	12	12	6	7	115 (15)	3235
<i>TRNA-TRP</i>	3	2	3	3	2	-	-	3	1	2	19 (6)	447
<i>TRNA-CYS</i>	11	13	9	14	5	14	11	10	11	12	110 (12)	2596
<i>TRNA-TYR</i>	-	1	-	-	-	-	-	1	1	1	4 (1)	100
<i>COX1</i>	1	1	2	2	5	3	7	3	3	4	31 (2)	1175
<i>TRNA-LEU1</i>	5	1	9	3	1	6	5	3	1	2	36 (10)	786
<i>COX2</i>	2	6	3	4	14	7	5	5	2	4	52 (9)	1492
<i>TRNA-LYS</i>	-	2	3	-	2	-	2	2	-	1	12 (3)	290
<i>TRNA-ASP</i>	3	-	1	2	-	2	1	3	-	-	12 (1)	273
<i>ATP8</i>	-	5	1	2	-	1	3	2	1	-	15 (3)	438
<i>ATP6</i>	4	7	4	4	5	2	2	1	2	3	34 (2)	957
<i>COX3</i>	3	2	2	-	4	2	3	-	3	4	23 (3)	717
<i>TRNA-GLY</i>	2	3	4	-	2	1	5	2	2	2	23 (3)	512
<i>ND3</i>	17	17	14	6	9	14	10	9	15	2	113 (21)	2876
<i>TRNA-ALA</i>	1	6	1	5	1	1	1	2	-	1	19 (0)	407
<i>TRNA-ARG</i>	2	1	5	5	2	5	-	1	1	3	25 (4)	695
<i>TRNA-ASN</i>	1	1	-	1	-	-	-	-	1	-	4 (1)	112
<i>TRNA-SER1</i>	-	-	-	-	-	-	-	2	-	-	2 (0)	52
<i>TRNA-GLU1</i>	5	7	4	2	4	3	4	4	5	-	38 (8)	924
<i>TRNA-PHE</i>	-	1	-	-	2	1	-	2	2	1	9 (2)	187
<i>ND5</i>	17	15	7	11	7	10	10	3	10	8	98 (10)	3244
<i>TRNA-HIS</i>	-	3	1	2	2	3	7	1	3	2	24 (6)	536
<i>ND4</i>	7	4	3	4	2	1	2	4	3	7	37 (1)	1194
<i>ND4L</i>	4	3	3	1	1	2	1	2	-	6	23 (2)	615
<i>TRNA-THR</i>	5	2	-	-	2	-	3	3	2	-	17 (9)	355

Table 2. Cont.

mtDNA Sequence	LG										Number of NUMTs (100% Identical) *	Sum of Gene Fragments (bp)
	1	2	3	4	5	6	7	8	9	10		
<i>TRNA-PRO</i>	-	-	1	3	-	-	1	-	1	-	6 (1)	151
<i>ND6</i>	9	14	12	7	5	8	14	10	4	8	91 (7)	2445
<i>CYTB</i>	8	14	7	5	6	5	4	3	6	4	62 (8)	1631
<i>TRNA-SER2</i>	9	7	9	3	9	6	5	4	5	11	68 (16)	2644
<i>ND1</i>	15	12	9	13	6	11	13	6	5	6	96 (16)	1486
<i>TRNA-LEU2</i>	-	3	-	-	1	1	-	1	-	-	6 (2)	135
<i>16S rRNA</i>	11	19	3	7	15	8	11	11	10	10	105 (10)	3700
<i>TRNA-VAL</i>	2	2	1	1	3	1	2	2	-	-	14 (4)	320
<i>12S rRNA</i>	12	16	6	7	5	6	12	5	13	10	92 (10)	2617
<i>TRNA-ILE</i>	-	-	-	-	-	-	-	-	-	-	0	0
<i>D-LOOP</i>	13	12	22	13	29	12	12	10	11	13	147 (23)	3979
<b>SUM</b>	<b>189</b>	<b>228</b>	<b>170</b>	<b>150</b>	<b>165</b>	<b>150</b>	<b>171</b>	<b>134</b>	<b>135</b>	<b>138</b>	<b>1630 (253)</b>	
<b>Sum of fragments in the LG (bp)</b>	5063	6071	4389	4147	4600	3997	4829	3533	3966	3796		44.391

\* The counts of NUMTs that were 100% identical with the sequence from mtDNA are indicated in brackets.

**Table 3.** Minimal and maximal length of each mtDNA gene fragments located in the nuclear genome.

Sequence	Length (in bp)	Min (in bp)	% Min	Max (in bp)	% Max
<i>TRNA-GLN</i>	68	16	24%	44	65%
<i>TRNA-MET</i>	65	16	25%	18	28%
<i>ND2</i>	987	18	2%	58	6%
<i>TRNA-TRP</i>	64	16	25%	38	59%
<i>TRNA-CYS</i>	64	16	25%	40	63%
<i>TRNA-TYR</i>	64	18	28%	35	55%
<i>COX1</i>	1537	19	1%	72	5%
<i>TRNA-LEU1</i>	68	16	24%	39	57%
<i>COX2</i>	679	18	3%	52	8%
<i>TRNA-LYS</i>	70	17	24%	35	50%
<i>TRNA-ASP</i>	63	17	27%	32	51%
<i>ATP8</i>	168	17	10%	54	32%
<i>ATP6</i>	675	18	3%	47	7%
<i>COX3</i>	789	19	2%	94	12%
<i>TRNA-GLY</i>	63	16	25%	32	51%
<i>ND3</i>	354	17	5%	99	28%
<i>TRNA-ALA</i>	62	19	31%	28	45%
<i>TRNA-ARG</i>	64	16	25%	61	95%
<i>TRNA-ASN</i>	67	16	24%	45	67%
<i>TRNA-SER1</i>	65	25	38%	27	42%
<i>TRNA-GLU</i>	65	16	25%	63	97%
<i>TRNA-PHE</i>	68	16	24%	26	38%
<i>ND5</i>	1708	19	1%	94	6%
<i>TRNA-HIS</i>	63	16	25%	42	67%
<i>ND4</i>	1315	19	1%	108	8%
<i>ND4L</i>	306	17	6%	38	12%
<i>TRNA-THR</i>	63	16	25%	60	95%
<i>TRNA-PRO</i>	64	16	25%	34	53%
<i>ND6</i>	504	18	4%	64	13%
<i>CYTB</i>	1133	18	2%	46	4%
<i>ND1</i>	927	18	2%	88	9%
<i>TRNA-SER2</i>	69	16	23%	35	51%
<i>TRNA-LEU2</i>	67	16	24%	27	40%
<i>16S rRNA</i>	1373	19	1%	78	6%
<i>TRNA-VAL</i>	72	16	22%	38	53%
<i>12S rRNA</i>	752	18	2%	72	10%
<i>TRNA-ILE</i>	64	-	-	-	-
<i>D-LOOP</i>	289	17	6%	182	63%

Of the 1630 NUMTs (Table 1), 253 fragments, representing 16% of all NUMTs, showed 100% homology with the mtDNA gene sequences. 100% sequence homology for *TRNA-MET* was found for all 6 NUMTs (Tables 2 and 4). 23 NUMTs whose sequence homology was 100% were observed for the D-loop region and 21 NUMTs for the gene *ND3*. In contrast, in the case of genes *TRNA-SER1* and *TRNA-ALA*, no NUMTs with 100% sequence homology were observed (Table 2). The mean values of the percentage of sequence identity ranged from 88.0% (*ATP8*) to 100% (*TRNA-MET*). The percentage identity for the individual linkage groups was in the range of 90–91%. The largest homology was recorded on LG10 (90.7%) and the lowest—on LG6 (90.2%) (Table 4). At least one sequence with 100% homology was identified on each of the linkage groups.

Table 5 compares the NUMTs' number and length with mtDNA fragments found in pseudogenes and transcriptome of *D. magna*. The overall number of mtDNA fragments in pseudogenes was 599 and their highest length was estimated between 18 bp (*TRNA-MET*) to 99 bp (*ND3*). The lowest number of mtDNA fragments being part of pseudogenes was observed for *TRNA-PRO* (1), *ATP8* (2), and *TRNA-ASP* and *TRNA-ILE* (3). The highest number of these fragments was observed for *16S rRNA* (63) and D-loop (48). In the case of mtDNA fragments found in the transcriptome, the highest number was observed for *TRNA-CYS* (101) and *16S rRNA* (100). The lowest number of mtDNA fragments was found for *TRNA-SER1* and *TRNA-PHE* gene (1). The highest length ranged between 20 bp to 64 bp. The main localization of mtDNA fragments in transcriptome was in different mRNA transcript variants of various protein-coding genes as well as ncRNA and misc\_RNA sequences (Table S1). The overall number of mtDNA fragments in transcriptome was 1275.

**Table 4.** Mean % identity of mtDNA gene fragments located in the linkage groups.

Sequence	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9	LG10	Mean % Identity–Gene
<i>TRNA-GLN</i>	93.8	88.8	94.3	92.9	92.3	85.1	97.1	89.4	90.5	95.0	<b>91.6</b>
<i>TRNA-MET</i>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	<b>100.0</b>	-	-	-	-	-	-	<b>100.0</b>
<i>ND2</i>	88.7	91.4	90.2	88.0	89.3	88.4	91.2	90.0	87.2	87.2	<b>89.5</b>
<i>TRNA-TRP</i>	90.9	95.2	88.3	90.2	93.8	-	-	89.7	<b>100.0</b>	94.0	<b>91.8</b>
<i>TRNA-CYS</i>	92.6	89.7	89.1	89.9	88.4	93.3	91.4	87.6	88.7	92.4	<b>90.5</b>
<i>TRNA-TYR</i>	-	85.7	-	-	-	-	-	<b>100.0</b>	94.7	97.1	<b>94.4</b>
<i>COX1</i>	79.5	92.3	82.6	94.6	87.8	82.2	88.3	90.7	92.8	95.4	<b>89.1</b>
<i>TRNA-LEU1</i>	91.0	<b>100.0</b>	92.5	91.9	<b>100.0</b>	93.6	91.5	95.2	86.2	83.3	<b>92.2</b>
<i>COX2</i>	81.6	95.4	86.5	87.5	92.1	90.2	91.1	89.0	85.0	89.0	<b>90.2</b>
<i>TRNA-LYS</i>	-	95.2	86.5	-	98.6	-	95.1	91.9	-	90.5	<b>92.6</b>
<i>TRNA-ASP</i>	96.8	-	95.0	84.4	-	85.6	90.9	91.5	-	-	<b>90.9</b>
<i>ATP8</i>	-	88.4	88.5	84.8	-	<b>100.0</b>	84.5	87.7	90.9	-	<b>88.0</b>
<i>ATP6</i>	86.9	85.1	90.4	88.4	93.1	90.3	86.9	95.2	85.5	92.7	<b>88.9</b>
<i>COX3</i>	84.9	90.5	87.6	-	90.5	91.3	97.1	-	88.3	94.6	<b>90.9</b>
<i>TRNA-GLY</i>	90.4	92.1	92.0	-	95.2	90.9	94.3	85.0	95.2	88.3	<b>92.0</b>
<i>ND3</i>	90.2	91.5	89.2	88.5	96.1	91.1	90.3	89.9	91.5	95.7	<b>91.0</b>
<i>TRNA-ALA</i>	90.5	93.7	94.7	89.8	94.7	90.9	88.0	90.7	-	94.7	<b>91.9</b>
<i>TRNA-ARG</i>	89.0	90.5	92.0	94.3	89.7	87.0	-	90.9	94.7	97.7	<b>91.7</b>
<i>TRNA-ASN</i>	84.6	<b>100.0</b>	-	88.0	-	-	-	-	95.6	-	<b>92.0</b>
<i>TRNA-SER1</i>	-	-	-	-	-	-	-	90.3	-	-	<b>90.3</b>
<i>TRNA-GLU</i>	89.8	90.8	89.3	90.2	93.0	95.1	87.6	91.6	96.2	-	<b>91.5</b>
<i>TRNA-PHE</i>	-	<b>100.0</b>	-	-	90.1	90.5	-	95.5	92.6	91.3	<b>93.1</b>
<i>ND5</i>	89.5	88.0	86.1	92.5	89.4	92.4	86.8	86.0	89.5	88.6	<b>89.2</b>
<i>TRNA-HIS</i>	-	95.1	<b>100.0</b>	<b>100.0</b>	80.7	87.6	91.2	88.9	89.0	<b>100.0</b>	<b>91.8</b>
<i>ND4</i>	89.5	86.1	90.9	87.3	93.0	89.7	87.3	88.5	88.7	94.1	<b>89.8</b>
<i>ND4L</i>	86.1	87.4	84.6	82.9	<b>100.0</b>	88.9	95.0	87.3	-	90.8	<b>88.5</b>
<i>TRNA-THR</i>	98.3	89.5	-	-	95.0	-	98.3	<b>100.0</b>	94.0	-	<b>96.6</b>
<i>TRNA-PRO</i>	-	-	91.3	90.7	-	-	<b>100.0</b>	-	92.6	-	<b>92.7</b>
<i>ND6</i>	92.9	89.0	89.7	92.6	87.8	90.9	89.1	89.2	88.7	92.6	<b>90.2</b>
<i>CYTB</i>	91.1	89.4	89.8	93.0	91.1	87.1	88.9	93.4	92.2	90.9	<b>90.5</b>
<i>ND1</i>	90.7	92.7	93.1	88.3	92.3	88.3	91.5	92.5	90.6	86.1	<b>90.6</b>

Table 4. Cont.

Sequence	LG1	LG2	LG3	LG4	LG5	LG6	LG7	LG8	LG9	LG10	Mean % Identity–Gene
<i>TRNA-SER2</i>	95.3	92.3	94.4	95.2	89.4	93.6	96.6	90.2	90.3	88.8	<b>92.3</b>
<i>TRNA-LEU2</i>	-	92.0	-	-	85.2	85.2	-	<b>100.0</b>	-	-	<b>91.1</b>
<i>16S RRNA</i>	86.3	87.6	81.7	91.9	87.1	88.4	86.4	87.9	86.1	92.7	<b>87.8</b>
<i>TRNA-VAL</i>	<b>100.0</b>	83.5	90.5	92.0	90.9	90.5	83.3	<b>100.0</b>	-	-	<b>91.4</b>
<i>12S RRNA</i>	89.1	89.9	88.2	87.8	90.2	88.9	89.2	93.0	93.1	85.5	<b>89.5</b>
<b>D-LOOP</b>	91.9	90.0	91.7	91.8	88.8	91.0	90.3	92.6	91.5	88.0	<b>90.6</b>
<b>Mean % Identity–LG</b>	<b>90.6</b>	<b>90.3</b>	<b>90.3</b>	<b>90.5</b>	<b>90.4</b>	<b>90.2</b>	<b>90.4</b>	<b>90.6</b>	<b>90.6</b>	<b>90.7</b>	<b>90.4</b>



**Table 5.** The comparison of the number and length of mitochondrial DNA fragments found in the nuclear genome (NUMTs), pseudogenes, and the transcriptome.

Gene	NUMTs		mtDNA Fragments Found in Pseudogenes		mtDNA Fragments Found in Transcriptome	
	Number	Greatest Length (bp)	Number	Greatest Length (bp)	Number	Greatest Length (bp)
<i>TRNA-GLN</i>	46	44	13	19	23	35
<i>TRNA-MET</i>	6	18	5	18	21	29
<i>ND2</i>	128	58	47	39	36	58
<i>TRNA-TRP</i>	23	38	11	30	58	38
<i>TRNA-CYS</i>	136	40	23	26	101	34
<i>TRNA-TYR</i>	4	35	0	0	5	25
<i>COX1</i>	37	72	11	72	52	61
<i>TRNA-LEU1</i>	38	39	5	20	28	23
<i>COX2</i>	70	53	21	38	14	39
<i>TRNA-LYS</i>	14	35	8	35	7	31
<i>TRNA-ASP</i>	14	32	3	29	18	39
<i>ATP8</i>	18	54	2	15	10	26
<i>ATP6</i>	38	64	11	24	6	64
<i>COX3</i>	29	94	15	27	25	28
<i>TRNA-GLY</i>	23	32	13	37	31	48
<i>ND3</i>	131	99	28	99	27	53
<i>TRNA-ALA</i>	21	28	5	18	57	28
<i>TRNA-ARG</i>	29	61	18	31	85	61
<i>TRNA-ASN</i>	4	45	4	19	3	45
<i>TRNA-SER1</i>	3	27	0	0	1	20
<i>TRNA-GLU</i>	46	63	15	29	27	30
<i>TRNA-PHE</i>	10	26	5	24	1	26
<i>ND5</i>	107	94	19	65	70	43
<i>TRNA-HIS</i>	26	42	7	21	28	27
<i>ND4</i>	46	108	25	33	73	33
<i>ND4L</i>	25	38	10	38	5	30
<i>TRNA-THR</i>	18	60	9	22	46	60
<i>TRNA-PRO</i>	6	34	1	13	19	38
<i>ND6</i>	106	64	21	41	97	42
<i>CYTB</i>	71	59	27	38	19	35
<i>TRNA-SER2</i>	74	35	29	37	33	31

Table 5. Cont.

Gene	NUMTs		mtDNA Fragments Found in Pseudogenes		mtDNA Fragments Found in Transcriptome	
	Number	Greatest Length (bp)	Number	Greatest Length (bp)	Number	Greatest Length (bp)
<i>ND1</i>	107	88	45	88	24	34
<i>TRNA-LEU2</i>	7	27	10	23	9	27
<i>16S rRNA</i>	123	78	63	59	100	46
<i>TRNA-VAL</i>	19	38	9	20	8	32
<i>12S rRNA</i>	136	72	10	25	42	51
<i>TRNA-ILE</i>	0	0	3	20	2	20
D-LOOP	170	182	48	34	64	50
Sum	1909		599		1275	

### 3. Discussion

This is the first study in which the *D. magna* nuclear genome deposited in the GenBank database was analyzed for pseudogene sequences of mitochondrial origin. The first complete information about the genome of *D. magna* was published by Lee et al. in 2019 [14]. To date, there is no information about pseudogenes localized in the genome of the water flea. Our research provides comprehensive bioinformatic details about the location of NUMTs found in the reference genome, which may be useful for future phylogenetic, evolution, and/or population analyses.

A computer-based search for NUMTs in the nuclear genome of 85 species of animals, plants, fungi, and protists showed that the total length of detected NUMTs varied from 0 to 823.9 kb per nuclear genome [23]. For instance, the NUMT content was 0 in *Anopheles gambiae*, 263.478 bp in *Homo sapiens*, and more than 800 kbp in *Oryza sativa* [23]. In the case of *D. magna* (Table 1), the overall length of NUMT in the nuclear genome was 44.391 bp. The total length of 24 NUMTs was 9.989 bp in the *Pteromalus puparum* genome, and 42.972 bp in *Nasoni vitripennis* [24], and more than 230 kbp in *Apis mellifera* [25]. There seems to be a positive correlation between the haploid genome sizes (C-values) and NUMT amount/prevalence in eukaryotes [26].

Another explanation for the differences in the total length of NUMTs is the fact that they accumulate in the genomes in a continuous evolutionary process [18]. Like in other species, e.g., *Myotis lucifugus* [27], the percentage of NUMTs in the genome was less than 0.1%. In contrast, the number and length of NUMTs may vary depending on the computer-based query for NUMTs in the BLAST tool, as in the case of the genome of *Canis lupus familiaris* [22,28]. In our study, the sequence search in BLASTN 2.6.0 implemented in CLC Genomics Workbench 12.0 yielded 1909 results, although 279 results were found in various scaffolds used during the sequencing of the *D. magna* genome. In this paper, however, the NUMT results from the scaffold were excluded due to the potential occurrence of artifacts created during sequencing, as observed by Shi et al. [27], where surprisingly, an entire mitochondrial genome was found in the scaffold AAPE02072785 in the *M. lucifugus* genome. It is also worth considering that, in this work, the BLAST search result took into account all search results, even those fragments whose length was only 16 bp (Table 3). The NUMTs found in scaffolds and mitochondrial DNA fragments found in pseudogenes and transcriptome are listed in supplementary materials (Table S1).

In the case of various mtDNA gene fragments, such as *TRNA-MET*, *COX1*, and *ND4* where their number is higher in the transcriptome than in the nuclear genome, it may be the result of the presence of multiple mRNA transcripts or ncRNA forms in the transcriptome. The variability of mRNA and ncRNA forms may be caused by alternative splicing. Moreover, the transcriptome is changing during the lifespan and may depend on the type of tissue [29]. It is not excluded that some NUMT are parts of protein-coding genes and/or participate in gene regulation through ncRNA formation. In this research, the authors focused on describing the presence of mitochondrial DNA in the nuclear genome, while the investigation of the role of NUMTs in gene expression and regulation in *D. magna* should be verified experimentally in further analyses.

NUMTs were used to define characteristics and to clarify phylogenetic inconsistencies suggested by paralog sequences [18,30]. The analysis performed by Mishmar et al. [31] revealed that mtDNA fragments, which were integrated into the nucleus before the radiation of modern human mtDNAs, confirming that mtDNAs similar to today's African macro-haplogroup L were the first human mtDNAs. The analysis of NUMTs in *D. magna* revealed that the latest evolutionary sequences are pseudogenes derived from the sequence *TRNA-MET* as the homology of the entire gene sequence was 100%. The homology of *TRNA-MET* gene fragments localized in transcriptome is below 100%, however, it could be a result of gene modifications after expression (Table S1). In addition, all mtDNA fragments in transcriptome might differ from each other due to different ways of modulating gene expression depending on the tissue. The lowest homology (70.33%) was characteristic for the pseudogene derived from the mtDNA D-loop sequence located on LG5 (Table S1). It may probably be the oldest element of mitochondrial DNA incorporated into the nuclear genome. However, due to the high

degree of mutation in the D-loop, the thesis requires further verification. Similarly, NUMTs derived from *TRNA-ALA* and *TRNA-SER1* may have been one of the first sequences derived from mtDNA. However, by assessing only sequence homology, the sequence of incorporation of mitochondrial pseudogenes into the nuclear genome cannot be determined. Nevertheless, as observed by Mishmar et al. [31], the nuclear genome accumulates mutational changes at a much slower rate than mtDNA. Hence, the sequences of “recent” NUMTs can provide valuable information about the mtDNA sequences of the earliest humans.

NUMTs exhibit different degrees of homology to their mitochondrial counterparts. They are variable in size, evenly distributed within and among chromosomes, and, in some cases, they are highly rearranged and/or fragmented [32]. However, the mitochondrial chromosome’s size does not correlate with the NUMT frequency or size distribution [19]. The transfer to the nucleus can be influenced by mitochondria’s vulnerability to stress and other factors that may cause the escape of mtDNA to the cytoplasm [32]. Mutations in mtDNA may occur in the entire mitochondrial genome; however, they are most frequently detected in the hypervariable regions of D-loops [33–35].

The number of mutations as well as their incidence in the D-loop area may be related to the number of NUMTs occurring in the nuclear genome. The higher the mutation rate, the greater the likelihood of transfer of the D-loop fragment into the cytoplasm followed by its incorporation into the nuclear genome as a pseudogene. In the *D. magna* genome, the highest numbers of pseudogenes from the D-loop (147) were observed, and only 23 of them had 100% sequence homology (Table 1).

No pseudogenes derived from the *TRNA-ILE* gene sequence were observed in the nuclear genome (Table 1). However, the detailed search of *TRNA-ILE* fragments in pseudogenes revealed the presence of three of them in LOC116920659, LOC116923942, and LOC116927475 and two fragments in the transcriptome (Table S1). The difference between the number of NUMTs and pseudogenes and transcriptome for the *TRNA-ILE* gene might be the result of the searching method. Pseudogenes were generated by automated computational analysis using gene prediction method: Gnomon [36]. Gnomon is a gene prediction HMM-based program. The core algorithm is based on Genscan which uses a 3-periodic fifth-order Hidden Markov Model (HMM) for the coding propensity score and incorporates descriptions of the basic transcriptional, translational, and splicing signals, as well as length distributions and compositional features of exons, introns, and intergenic regions. For the genes for which no experimental information is available, Gnomon is creating conventional ab initio predictions [36]. As a result of this prediction, the proposed gene contains the fragment of *TRNA-ILE*. The presence of this sequence fragments derived from mtDNA cannot be excluded, however, further experimental analysis of the protein should be held. Due to the lack of comparison with other genome sequences for *D. magna* the authors could not exclude any of the results. The parameters of the Gnomon prediction were widely used in the description of other model organisms’ genomes i.e., *Arabidopsis thaliana*, *Danio rerio*, *Mus musculus*, and *Drosophila melanogaster* [36]. Thus, the authors recognize the possibility of occurrence of tRNA-ILE-derived NUMTs in the *D. magna* genome.

In the human genome, NUMTs are commonly associated with repetitive elements, suggesting a possible role for transposable elements in mtDNA integration in the nuclear genome [31]. Certain NUMTs are repeated multiple times within the human genome [32,37]. In the case of the *D. magna* genome, some NUMTs were also observed, which were repeated many times in different linkage groups (Table S1). However, their association with repetitive elements in the nuclear genome requires additional research. The average levels of NUMT sequence homology for the individual linkage groups do not differ significantly from each other, which may indicate a random and even inclusion of sequence fragments into each of them (Table 4).

The cytochrome C oxidase subunit I (*COI*) has possibly been the most commonly studied marker. However, its popularity is mainly associated with its use as a marker for the DNA barcoding of animal diversity [38]. There are several factors causing inadequacy of mtDNA in general and *COI* individually, such as male-biased gene flow, the selection on any mtDNA nucleotide(s) (as the whole genome is one linkage group), retention of ancestral polymorphism, and introgression following hybridization [39]. Presently, there are huge numbers of *COI* sequences in public databases, and most of them have a

limited length, generally close to the length of the barcoding region. It is known that the possibility of the presence of NUMTs in the existing data should not be ignored [40]. Since the success of taxonomic differentiation is positively correlated with the barcode length, the minibarcode length is usually kept above 100 bp. For example, an approximately 250-bp region of 16S rRNA can be successfully amplified from various medicinal preparations and food products. It provides the correct identification of animal species [41,42]. Gene fragments that are often used for species identification in *D. magna* are in the following ranges: *COX1* (19–72 bp), *CYTB* (18–46 bp), 12s rRNA (18–72 bp), and 16s rRNA (19–78 bp); each of them constitutes less than 10% of the length of the entire gene (Table 3). Hence, the NUMT sequences of frequently analyzed genes are generally shorter than the respective mitochondrial sequence; thus, the possibility of NUMT co-amplification should decrease with an increased length of the targeted mitochondrial marker. However, it is worth paying attention to the coverage of NUMTs derived from the *TRNA-ARG* (95%), *TRNA-GLU* (97%), and *TRNA-THR* (95%) genes (Table 3). The transcriptome analysis for *TRNA-ARG* revealed the existence of the NUMT in lncRNA sequence transcript variants of the gene LOC116925827 (Table S1). The presence of this gene fragment in a non-coding sequence may suggest its role in gene regulation. The *TRNA-THR*-derived NUMT was identified in mRNA gene transcript variant LOC116930084. On the other hand, the sequence fragment of *TRNA-GLU* was only partially found in the transcriptome in zinc finger CCHC domain-containing protein 7-like mRNA localized in LG2. Perhaps, in these cases, they are not NUMTs but functional gene fragments taking part in gene modulation. Nonetheless, to confirm it, a detailed analysis of the transcriptome on larger research groups should be held.

NUMTs are highly polymorphic in terms of the sequence, homo/heterozygosity status, and presence/absence at a specific locus [43]. These features facilitate the use of NUMTs as specific population markers, as proposed for the human population by Lang et al. [44]. The biological importance of NUMTs may correlate with their location on the chromosome. Depending on the insertion location, NUMTs may perturb the function of the genes [45]. Additionally, de novo integration of NUMT pseudogenes into the nuclear genome has an adverse effect in some cases: promoting various disorders and aging, as observed in humans [46]. Chatre and Ricchetti [47] report that migratory mitochondrial DNA can also impact the replication of the nuclear region in *Saccharomyces cerevisiae*.

#### 4. Materials and Methods

The whole sequence of the *Daphnia magna* genome, strain SK and annotation of the nuclear and mitochondrial genome of *D. magna* were obtained from GenBank (the accession numbers for the nuclear and mitochondrial genome are GCA\_003990815.1 and NC\_026914.1, respectively). Each of the mitochondrial genes were aligned separately to detect plausible NUMTs in the nuclear genome. The presence of mtDNA fragments in defined pseudogenes was estimated by the alignment of mtDNA sequences with the GCF\_003990815.1\_ASM399081v1\_pseudogene\_without\_product.fna sequence. The presence of mtDNA fragments in the transcriptome was established by the search in GCF\_003990815.1\_ASM399081v1\_rna.fna sequence. The detailed results were presented in Supplementary Table S1. The SK strain genome assembly was released on 02-Jan-2019 and was integrated with the high-density genetic linkage map of the strain Xinb3. The presence of NUMTs in the *D. magna* nuclear genome GCA\_003990815.1 was evaluated using the BLAST (BLASTN 2.6.0) (Table S1) [48] program implemented within the CLC Genomics Workbench 12.0 software package (<https://www.qiagenbioinformatics.com>). In the presented study, we implemented a low complexity filter to mask off the query sequence segments with low compositional complexity. Such filtering was used to eliminate statistically significant but biologically uninteresting reports from the BLAST output, leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences. The expectation value (e-value = 10) describing the threshold for reporting matches against database sequences based on the assumption that ten matches are expected to be found merely by chance according to the stochastic model of Karlin and Altschul (1990) [49]. If the E-value ascribed to a match was greater than the expected threshold, no match was reported. The lower

value of the threshold caused more stringent searching criteria, leading to fewer chance matches being reported. Higher threshold results in more matches being reported, but many may just match by chance, not due to any biological similarity. The value of match/mismatch, which assigned a score for aligning and evaluated the quality of a pairwise sequence alignment was set to 2/3. The penalty to open gap and penalty to extend gap (Gap Cost.) was set to 5 for existence and 2 for extension. The maximum number of database sequences, where BLAST found matches to a query sequence, to be included in the BLAST report was set to 100.5.

## 5. Conclusions

This article described the first occurrence of mitochondrial pseudogenetic sequences (NUMTs) in the nuclear genome of *D. magna*. There was no full sequence homology for two genes: *TRNA-SER1* as well as *TRNA-ALA* NUMTs. The total length of NUMTs in the nuclear genome was 44,391 bp (from 16 to 182 bp), which accounted for 0.042% of the entire genome. The best-matched NUMTs covering more than 90% of the mtDNA gene sequence were identified for the *TRNA-ARG* (95%), and *TRNA-THR* (95%) genes, and they may be included in the functional nuclear RNA molecules. The NUMT length varied from 16 to 63 bp for tRNA genes, from 17 to 108 bp for coding genes, from 18 to 78 bp for rRNA genes, and from 17 to 182 bp for the D-loop region. Therefore, using the product of total DNA isolation in mtDNA studies, co-amplification of nDNA fragments is unlikely especially in the case of amplification of the whole tRNA genes and fragments of other genes the D-loop with a length exceeding 200 bp. It was observed that fragments *TRNA-MET* (from 16 to 18 bp length) had the highest level of sequence homology, which means that they could be evolutionarily the youngest. The lowest degree of homology was found in the pseudogene derived from the mtDNA D-loop sequence. It may probably be the oldest element of mitochondrial DNA incorporated into the nuclear genome; however, due to the high degree of mutation in the D-loop, the thesis requires further analysis and elucidation.

**Supplementary Materials:** Supplementary materials can be found at <http://www.mdpi.com/1422-0067/21/22/8725/s1>. Table S1: The presence and evaluation of all NUMTs found in the BLAST tool as well as mtDNA fragments found in pseudogenes and transcriptome.

**Author Contributions:** Conceptualization, K.K. and B.Ś.; methodology, K.K. and M.P.; software, M.P.; validation, B.Ś., and A.B.; formal analysis, K.K.; investigation, M.P.; writing—original draft preparation, K.K.; writing—review and editing, A.T. and B.Ś.; project administration, B.Ś. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

NUMTs	nuclear copies of mitochondrial dna
BP	base pairs
LG	linkage groups
KIT	Korea Institute of Toxicology
EST	expressed sequence tags
nDNA	nuclear DNA
mtDNA	mitochondrial DNA
BLAST	Basic Local Alignment Search Tool
ND1-6 and ND4L	NADH dehydrogenase subunits 1–6 and subunit 4L
CO1-3 or COX1-3	cytochrome oxidase subunits I-III
ATP6 and 8	ATPase subunit 6 and 8
CYTB	cytochrome b
12S rRNA	gene for small subunit ribosomal RNA
16S rRNA	gene for large subunit ribosomal RNA
TRNA	gene coding transfer RNA



## References

1. Ebert, D. A genome for the environment. *Science* **2011**, *331*, 539–540. [[CrossRef](#)] [[PubMed](#)]
2. Lampert, W. Daphnia: Development of a model organism in ecology and evolution. *Excell. Ecol.* **2011**, *21*, 1–275.
3. Miner, B.E.; de Meester, L.; Pfrender, M.E.; Lampert, W.; Hairston, N.G. Linking genes to communities and ecosystems: Daphnia as an ecogenomic model. *Proc. R. Soc. B Biol. Sci.* **2012**, *279*, 1873–1882. [[CrossRef](#)] [[PubMed](#)]
4. Bownik, A.; Ślaska, B.; Bochra, J.; Gumieniak, K.; Gałek, K. Procaine penicillin alters swimming behaviour and physiological parameters of *Daphnia magna*. *Environ. Sci. Pollut. Res.* **2019**, *26*, 18662–18673. [[CrossRef](#)] [[PubMed](#)]
5. Bownik, A.; Jasieczek, M.; Kosztowny, E. Ketoprofen affects swimming behavior and impairs physiological endpoints of *Daphnia magna*. *Sci. Total Environ.* **2020**, 725. [[CrossRef](#)] [[PubMed](#)]
6. Bownik, A.; Szabelak, A.; Kulińska, M.; Wałęka, M. Effects of L-proline on swimming parameters of *Daphnia magna* subjected to heat stress. *J. Therm. Biol.* **2019**, *84*, 154–163. [[CrossRef](#)]
7. Tkaczyk, A.; Bownik, A.; Dudka, J.; Kowal, K.; Ślaska, B. *Daphnia magna* model in the toxicity assessment of pharmaceuticals: A review. *Sci. Total Environ.* **2020**, in press. [[CrossRef](#)]
8. Bownik, A.; Ślaska, B.; Dudka, J. Cisplatin affects locomotor activity and physiological endpoints of *Daphnia magna*. *J. Hazard. Mater.* **2020**, 384. [[CrossRef](#)]
9. Persoone, G.; Baudo, R.; Cotman, M.; Blaise, C.; Thompson, K.C.; Moreira-Santos, M.; Vولات, B.; Törökne, A.; Han, T. Review on the acute *Daphnia magna* toxicity test? Evaluation of the sensitivity and the precision of assays performed with organisms from laboratory cultures or hatched from dormant eggs. *Knowl. Manag. Aquat. Ecosyst.* **2009**, 1–29. [[CrossRef](#)]
10. OECD. Daphnia Acute Immobilisation Test and Reproduction Test. In *OECD Guideline for Testing of Chemicals*; 1984; pp. 1–16. Available online: <https://www.oecd.org/chemicalsafety/risk-assessment/1948249.pdf> (accessed on 18 November 2020).
11. Routtu, J.; Hall, M.D.; Albere, B.; Beisel, C.; Bergeron, R.D.; Chaturvedi, A.; Choi, J.H.; Colbourne, J.; Meester, L.D.; Stephens, M.T.; et al. An SNP-based second-generation genetic map of *Daphnia magna* and its application to QTL analysis of phenotypic traits. *BMC Genom.* **2014**, *15*, 1–15. [[CrossRef](#)]
12. Routtu, J.; Jansen, B.; Colson, I.; De Meester, L.; Ebert, D. The first-generation *Daphnia magna* linkage map. *BMC Genomics* **2010**, *11*. [[CrossRef](#)] [[PubMed](#)]
13. Dukić, M.; Berner, D.; Roesti, M.; Haag, C.R.; Ebert, D. A high-density genetic map reveals variation in recombination rate across the genome of *Daphnia magna*. *BMC Genet.* **2016**, *17*, 1–13. [[CrossRef](#)] [[PubMed](#)]
14. Lee, B.Y.; Choi, B.S.; Kim, M.S.; Park, J.C.; Jeong, C.B.; Han, J.; Lee, J.S. The genome of the freshwater water flea *Daphnia magna*: A potential use for freshwater molecular ecotoxicology. *Aquat. Toxicol.* **2019**, *210*, 69–84. [[CrossRef](#)] [[PubMed](#)]
15. Colson, I.; Du Pasquier, L.; Ebert, D. Intragenic tandem repeats in *Daphnia magna*: Structure, function and distribution. *BMC Res. Notes* **2009**, *2*. [[CrossRef](#)] [[PubMed](#)]
16. Lopez, J.V.; Yuhki, N.; Masuda, R.; Modi, W.; O'Brien, S.J. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. *J. Mol. Evol.* **1994**, *39*, 174–190. [[CrossRef](#)] [[PubMed](#)]
17. Sun, X.; Yang, A. Exceptionally large mitochondrial fragments to the nucleus in sequenced mollusk genomes. *Mitochondrial DNA Part A* **2016**, *27*, 1409–1410. [[CrossRef](#)]
18. Song, H.; Moulton, M.J.; Hiatt, K.D.; Whiting, M.F. Uncovering historical signature of mitochondrial DNA hidden in the nuclear genome: The biogeography of *Schistocerca* revisited. *Cladistics* **2013**, *29*, 643–662. [[CrossRef](#)]
19. Leister, D. Origin, evolution and genetic effects of nuclear insertions of organelle DNA. *Trends Genet.* **2005**, *21*, 655–663. [[CrossRef](#)]
20. Williams, S.T.; Knowlton, N. Mitochondrial pseudogenes are pervasive and often insidious in the snapping shrimp genus *Alpheus*. *Mol. Biol. Evol.* **2001**, *18*, 1484–1493. [[CrossRef](#)]
21. Bensasson, D.; Zhang, D.X.; Hartl, D.L.; Hewitt, G.M. Mitochondrial pseudogenes: Evolution's misplaced witnesses. *Trends Ecol. Evol.* **2001**, *16*, 314–321. [[CrossRef](#)]

22. Verscheure, S.; Backeljau, T.; Desmyter, S. In silico discovery of a nearly complete mitochondrial genome Numt in the dog (*Canis lupus familiaris*) nuclear genome. *Genetica* **2015**, *143*, 453–458. [[CrossRef](#)] [[PubMed](#)]
23. Hazkani-Covo, E.; Zeller, R.M.; Martin, W. Molecular poltergeists: Mitochondrial DNA copies (numts) in sequenced nuclear genomes. *PLoS Genet.* **2010**, *6*. [[CrossRef](#)] [[PubMed](#)]
24. Viljakainen, L.; Oliveira, D.C.S.G.; Werren, J.H.; Behura, S.K. Transfers of mitochondrial DNA to the nuclear genome in the wasp *Nasonia vitripennis*. *Insect Mol. Biol.* **2010**, *19*, 27–35. [[CrossRef](#)] [[PubMed](#)]
25. Behura, S.K. Analysis of nuclear copies of mitochondrial sequences in honeybee (*Apis mellifera*) genome. *Mol. Biol. Evol.* **2007**, *24*, 1492–1505. [[CrossRef](#)] [[PubMed](#)]
26. Song, H.; Moulton, M.J.; Whiting, M.F. Rampant nuclear insertion of mtDNA across diverse lineages with in Orthoptera (Insecta). *PLoS ONE* **2014**, *9*, 41–43. [[CrossRef](#)] [[PubMed](#)]
27. Shi, H.; Xing, Y.; Mao, X. The little brown bat nuclear genome contains an entire mitochondrial genome: Real or artifact? *Gene* **2017**, *629*, 64–67. [[CrossRef](#)] [[PubMed](#)]
28. Ishiguro, N.; Nakajima, A.; Horiuchi, M.; Shinagawa, M. Multiple nuclear pseudogenes of mitochondrial DNA exist in the canine genome. *Mamm. Genome* **2002**, *13*, 365–372. [[CrossRef](#)]
29. Lai, R.W.; Lu, R.; Danthi, P.S.; Bravo, J.I.; Goumba, A.; Sampathkumar, N.K.; Benayoun, B.A. Multi-level remodeling of transcriptional landscapes in aging and longevity. *BMB Rep.* **2019**, *52*, 86–108. [[CrossRef](#)]
30. Berthier, K.; Chapuis, M.P.; Moosavi, S.M.; Tohidi-Esfahani, D.; Sword, G.A. Nuclear insertions and heteroplasmy of mitochondrial DNA as two sources of intra-individual genomic variation in grasshoppers. *Syst. Entomol.* **2011**, *36*, 285–299. [[CrossRef](#)]
31. Mishmar, D.; Ruiz-Pesini, E.; Brandon, M.; Wallace, D.C. Mitochondrial DNA-Like Sequences in the Nucleus (NUMTs): Insights into Our African Origins and the Mechanism of Foreign DNA Integration. *Hum. Mutat.* **2004**, *23*, 125–133. [[CrossRef](#)]
32. Woischnik, M.; Moraes, C.T. Pattern of Organization of Human Mitochondrial Pseudogenes in the Nuclear Genome. *Genome Res.* **2002**, *12*, 885–893. [[CrossRef](#)] [[PubMed](#)]
33. Bertagnolli, A.C.; Soares, P.; van Asch, B.; Amorim, A.; Cirnes, L.; Máximo, V.; Cassali, G.D. An assessment of the clonality of the components of canine mixed mammary tumours by mitochondrial DNA analysis. *Vet. J.* **2009**, *182*, 269–274. [[CrossRef](#)] [[PubMed](#)]
34. Murgia, C.; Pritchard, J.K.; Kim, S.Y.; Fassati, A.; Weiss, R.A. Clonal Origin and Evolution of a Transmissible Cancer. *Cell* **2006**, *126*, 477–487. [[CrossRef](#)] [[PubMed](#)]
35. Ślaska, B.; Grzybowska-Szatkowska, L.; Bugno-Poniewierska, M.; Surdyka, M.; Śmiech, A. Nuclear and mitochondrial DNA mutation in human and canine tumors. *Med. Weter.* **2013**, *69*, 195–202.
36. Souvorov, A.; Kapustin, Y.; Kiryutin, B.; Chetvernin, V.; Tatusova, T.; Lipman, D. *Gnomon-NCBI Eukaryotic Gene Prediction Tool*; National Center for Biotechnology Information: Bethesda, MD, USA, 2010.
37. Tourmen, Y.; Baris, O.; Dessen, P.; Jacques, C.; Malthiery, Y.; Reynier, P. Structure and Chromosomal Distribution of Human Mitochondrial Pseudogenes. *Genomics* **2002**, *80*, 71–77. [[CrossRef](#)]
38. Hebert, P.D.N.; Dewaard, J.R.; Landry, J.F. DNA barcodes for 1/1000 of the animal Kingdom. *Biol. Lett.* **2010**, *6*, 359–362. [[CrossRef](#)]
39. Moritz, C.; Cicero, C. DNA barcoding: Promise and pitfalls. *PLoS Biol.* **2004**, *2*. [[CrossRef](#)]
40. Bertheau, C.; Schuler, H.; Krumböck, S.; Arthofer, W.; Stauffer, C. Hit or miss in phylogeographic analyses: The case of the cryptic NUMTs. *Mol. Ecol. Resour.* **2011**, *11*, 1056–1059. [[CrossRef](#)]
41. Arulandhu, A.J.; Staats, M.; Hagelaar, R.; Voorhuijzen, M.M.; Prins, T.W.; Scholtens, I.; Costessi, A.; Duijsings, D.; Rechenmann, F.; Gaspar, F.B.; et al. Development and validation of a multi-locus DNA metabarcoding method to identify endangered species in complex samples. *Gigascience* **2017**, *6*, 1–18. [[CrossRef](#)]
42. Coghlan, M.L.; Haile, J.; Houston, J.; Murray, D.C.; White, N.E.; Moolhuijzen, P.; Bellgard, M.I.; Bunce, M. Deep sequencing of plant and animal DNA contained within traditional Chinese medicines reveals legality issues and health safety concerns. *PLoS Genet.* **2012**, *8*. [[CrossRef](#)]
43. Hazkani-Covo, E.; Graur, D. A comparative analysis of numt evolution in human and chimpanzee. *Mol. Biol. Evol.* **2007**, *24*, 13–18. [[CrossRef](#)] [[PubMed](#)]
44. Lang, M.; Sazzini, M.; Calabrese, F.M.; Simone, D.; Boattini, A.; Romeo, G.; Luiselli, D.; Attimonelli, M.; Gasparre, G. Polymorphic NumtS trace human population relationships. *Hum. Genet.* **2012**, *131*, 757–771. [[CrossRef](#)] [[PubMed](#)]



45. Gaziev, A.I.; Shaikhaev, G.O. Nuclear mitochondrial Pseudogenes. *Mol. Biol.* **2010**, *44*, 358–368. [[CrossRef](#)]
46. Dayama, G.; Emery, S.B.; Kidd, J.M.; Mills, R.E. The genomic landscape of polymorphic human nuclear mitochondrial insertions. *Nucleic Acids Res.* **2014**, *42*, 12640–12649. [[CrossRef](#)] [[PubMed](#)]
47. Chatre, L.; Ricchetti, M. Nuclear mitochondrial DNA activates replication in *Saccharomyces cerevisiae*. *PLoS ONE* **2011**, *6*. [[CrossRef](#)]
48. Roesch, L.F.W.; Fulthorpe, R.R.; Jacques, R.J.S.; Bento, F.M.; de Oliveira Camargo, F.A. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *World J. Microbiol. Biotechnol.* **2006**, *22*, 3389–3402. [[CrossRef](#)]
49. Karlin, S.; Altschul, S.F. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 2264–2268. [[CrossRef](#)]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).