Available online at www.sciencedirect.com
SCIENCE DIRECT
Gene 354 (2005) 181 – 188
GENE
SECTION
FUNCTIONAL GENOMICS
www.elsevier.com/locate/gene

Evolution of nuclearly encoded mitochondrial genes in Metazoa

Anna De Grassi[a], Corrado Caggese[b], Domenica D'Elia[a], Cecilia Lanave[a], Graziano Pesole[c], Cecilia Saccone[a,d,*]

[a]*Istituto di Tecnologie Biomediche, Sezione di Bari, CNR, Bari, Italy*
[b]*Dipartimento di Anatomia Patologica e di Genetica (DAPEG), sezione di Genetica, Universita` di Bari, Bari, Italy*
[c]*Dipartimento di Scienze Biomolecolari e Biotecnologie, Università degli Studi di Milano, Milano, Italy*
[d]*Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Bari, Italy*

Received 15 December 2004; received in revised form 25 February 2005; accepted 24 March 2005
Available online 21 June 2005
Received by W. Gray

Abstract

All Metazoan nuclear genomes underwent a continuous process of both complete and partial genetic material gain and loss. The forces modulating these events are also subject to the strict interaction between nuclear and mitochondrial (mt) genome. In this context we investigate the evolution of nuclear genes encoding proteins which target the mitochondrion, with a particular attention to genes involved in oxidative phosphorylation (OXPHOS), one of the most ancient and conserved functions. To examine thoroughly the evolutionary strategies that preserve OXPHOS and coordinate the two cellular genomes, a comparative analysis has been carried out for 78 OXPHOS gene families in several Metazoa (insects, tunicates, fishes and mammals). We demonstrate that the duplication rate of OXPHOS genes increases passing from invertebrates to vertebrates consistently with the total increase in genome size, but all species are prone to negatively select OXPHOS duplicates compared to the general trend of nuclear gene families. These results are consistent with the 'balance hypothesis' and, at least in insects, the expression of duplicate genes is low and strongly testis-biased.

Keywords: Mitochondrion; Oxidative phosphorylation; Gene families

1. Introduction

It is well-known that all eukaryotes, except few organisms like amitochondriate protists, possess more than one genome. In respiring cells, in addition to the nuclear genome, there is the mt genome, which has a reduced size. It derives, according to the most popular and accepted theory (one step endosymbiontic theory), from a primitive eubacterium which, together with an archaebacterium, constituted the primordial eukaryotic cell (Margulis, 1970; Martin and Muller, 1998).

During evolution, most of the genes belonging to the eubacterium from which mitochondrion originated was lost or transferred to the nucleus, which then became a chimera possessing genetic material from both the eu—and the archaebacterium.

In all extant eukaryotes, mitochondrial components are mainly coded by nuclear genomes and, only to a very small extent, by mt genome.

In Metazoa the transfer of genetic material from the mitochondrion to the nucleus has apparently stopped and thus the mt genome is practically in a "frozen state" in the ~800 MY of Metazoan evolution. The contribution of mt DNA is restricted to 13 proteins, 2 ribosomal RNAs and 22 tRNAs with few small exceptions. Size and shape of mt DNA are also almost frozen but the plasticity of the genome is kept through several change events, the most important ones consisting in: nucleotide changes, small insertions and

Abbreviations: Mt, Mitochondrial; OXPHOS, Oxidative phosphorylation; MCL, Markov cluster; Myr, Million years.

* Corresponding author. Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, Bari, Italy. Tel.: +39 080 5929661; fax: +39 080 5929690.
E-mail address: cecilia.saccone@ba.itb.cnr.it (C. Saccone).

0378-1119/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2005.03.046

Table 1
Gene family size of 78 OXPHOS genes

Complex	Human SwissProt ID	ENSEMBL family description	Insects		Tunicates		Fishes		Mammals			
			Ag	Dm	Ci	Cs	Dr	Fr	Mm	Rn	Pt	Hs
I	O00217	23 kda subunit	1	1	1	1	2	2	1	1	1	1
I	O00483	MLRQ subunit	2	1	1	1	0	2	2	2	3	3
I	O14561	acyl carrier protein	1	1	1	1	3	1	1	2	1	1
I	O43181	18 kDa subunit	1	1	1	1	2	1	1	1	1	1
I	O43674	SGDH subunit	1	1	1	1	2	0	1	1	1	1
I	O43676	B12 subunit	1	1	0	0	0	1	1	2	2	2
I	O43678	B8 subunit	2	1	1	1	1	1	1	2	1	1
I	O43920	15 kDa subunit	0	0	1	1	1	1	2	3	3	3
I	O75251	20 kDa subunit	3	2	1	1	0	1	1	1	1	1
I	O75306	49 kDa subunit	2	2	1	1	1	1	1	1	2	1
I	O75380	13 kDa A subunit	1	1	1	1	1	1	2	3	1	1
I	O75438	Unknown	0	0	1	1	0	1	0	0	1	1
I	O75489	30 kDa subunit	1	1	1	1	1	1	2	1	1	1
I	O95139	B17 subunit	0	0	1	1	0	0	1	1	2	1
I	O95168	B15 subunit	0	0	1	0	2	1	2	8	2	3
I	O95169	ASHI subunit	2	1	1	1	1	1	1	1	1	1
I	O95178	AGGG subunit	1	1	1	0	1	0	1	1	1	1
I	O95182	subunit B14 5A	2	2	1	0	1	1	1	2	1	1
I	O95298	subunit B14 5B	0	0	1	0	0	1	1	1	1	1
I	O95299	42 kDa subunit	1	1	1	1	1	1	1	2	2	1
I	O96000	PDSW subunit	1	1	1	1	1	1	1	1	1	1
I	P17568	B18 subunit	1	1	1	1	1	1	1	1	1	1
I	P19404	24 kDa subunit	1	2	1	1	1	1	1	1	2	2
I	P28331	75 kDa subunit	1	1	1	1	1	2	1	1	1	1
I	P49821	51 kDa subunit	2	3	1	1	1	2	1	1	1	1
I	P51970	19 kDa subunit	1	1	1	1	2	1	1	1	1	1
I	P56556	B14 subunit	2	1	1	1	0	1	1	1	1	1
I	Q16718	13 kDa B subunit	1	1	1	1	0	1	1	2	2	2
I	Q16795	39 kDa subunit	1	1	1	1	2	1	1	1	2	1
I	Q9P0J0	B16 6 subunit	1	1	1	0	1	1	1	2	1	1
I	Q9UI09	B17 2	1	1	1	1	1	1	1	1	1	1
I	Q9Y6M9	B22 subunit	1	1	1	1	3	1	1	1	2	2
II	O14521	Cytochrome *b* small subunit	1	1	0	0	0	0	1	2	1	1
II	P21912	Iron sulfur protein	1	2	1	1	1	2	1	1	1	1
II	P31040	Flavoprotein subunit	1	2	1	1	2	1	1	1	6	4
II	Q99643	Cytochrome *b*560 subunit	1	2	1	1	0	2	1	1	2	3
III	O14949	Ubiquinone binding QP C	1	1	0	0	2	1	1	1	1	1
III	O14957	Unknown	0	0	0	0	0	1	2	2	1	1
III	P07919	11 kda protein	1	1	1	1	0	2	3	1	1	1
III	P08574	Cytochrome *c*1 heme protein	1	2	1	1	2	1	1	1	1	1
III	P14927	14 kDa subunit	1	2	1	1	1	1	2	4	1	2
III	P22695	Core 2	1	1	1	1	2	2	1	1	1	1
III	P31930	Mitochondrial processing peptidase subunit	5	2	1	1	3	3	3	4	3	3
III	P47985	Iron sulfur subunit	1	1	1	1	1	1	1	1	3	3
III	Q9UDW1	Fragment	1	1	1	1	0	1	2	2	1	1
IV	P09669	Unknown	0	0	1	0	0	1	3	3	2	2
IV	P10606	Polypeptide VB	1	2	1	1	3	2	2	2	1	1
IV	P13073	Subunit IV 1	1	2	1	1	3	1	2	2	2	2
IV	P14854	Polypeptide VIB	1	1	1	1	1	3	2	4	2	2
IV	P15954	Unknown	1	1	1	0	0	1	4	0	1	1
IV	P20674	Polypeptide VA	1	1	1	1	2	2	1	2	1	1
IV	P24310	Polypeptide VIIA	0	0	1	0	2	3	4	4	3	4
IV	Q02221	Polypeptide VIA	2	2	1	1	1	2	2	2	3	3
V	O75947	D chain	1	1	1	1	1	1	2	2	3	2
V	O75964	G chain	1	2	1	1	1	2	1	6	1	1
V	P05496	Lipid binding protein	3	1	1	1	3	3	5	6	4	4
V	P06576	Beta chain	5	2	1	1	1	1	1	2	1	1
V	P18859	Coupling factor 6	1	2	1	1	2	2	1	2	1	1
V	P24539	B chain	1	1	1	1	1	1	1	1	3	3
V	P25705	Alpha chain	1	1	1	1	1	2	1	1	2	2

Table 1 (*continued*)

Complex	Human SwissProt ID	ENSEMBL family description	Insects		Tunicates		Fishes		Mammals			
			Ag	Dm	Ci	Cs	Dr	Fr	Mm	Rn	Pt	Hs
V	P30049	Delta chain	1	1	1	1	2	1	1	1	2	1
V	P36542	Gamma chain	1	1	1	1	1	1	1	1	1	1
V	P48047	Oligomycin sensitivity conferral protein	1	1	1	1	1	1	1	1	1	1
V	P56134	F chain	0	0	1	1	0	2	3	2	4	5
V	P56381	Unknown	0	0	1	1	0	1	1	1	2	2
V	P56385	Unknown	0	0	1	1	2	1	1	1	1	1
Accessory	O75880	Homolog mitochondrial precursor	1	1	1	1	1	1	0	2	2	2
Accessory	P00001	Cytochrome *c*	1	2	1	1	1	3	6	6	4	5
Accessory	P13804	ETF alpha subunit	1	1	1	1	1	1	1	1	1	1
Accessory	P38117	ETF beta subunit	1	1	1	1	1	1	1	1	1	1
Accessory	Q12887	Heme O synthase	1	1	1	1	2	2	1	1	1	1
Accessory	Q14061	Unknown	1	1	1	1	0	0	1	1	1	1
Accessory	Q15070	OXA1	1	1	1	1	1	1	1	1	0	1
Accessory	Q15526	Surfeit locus	2	2	1	1	3	2	2	2	2	4
Accessory	Q16134	Ubiquinone oxidoreductase ETF	2	1	1	1	0	1	1	1	1	1
Accessory	Q99766	ATP synthase coupling factor B	1	1	1	1	1	1	1	1	1	1
Accessory	Q9Y375	Complex I intermediate associated 30	1	1	1	1	1	1	1	1	1	1
Accessory	Q9Y6N1	COX11	1	1	1	1	0	1	1	1	2	2

Rows represent each of the 78 OXPHOS gene families: 66 subunits of respiratory complexes (I, II, III, IV, V) and 12 mitochondrial proteins associated to respiratory complexes (accessory). The ten principal columns report the number of genes for the corresponding OXPHOS gene family in the indicated species (Ag: *A. gambiae,* Dm: *D. melanogaster,* Ci: *C. intestinalis,* Cs: *C. savignyi,* Dr: *D. rerio,* Fr: *F. rubripes,* Mm: *M. musculus,* Rn: *R. norvegicus,* Pt: *P. troglodytes,* Hs: *H. sapiens*).

deletions, different gene distribution along the molecule, the presence of a variable regulatory region. These evolutionary features have been extensively studied by our group and discussed in previous papers and reviews (Saccone, 1994; Saccone et al., 1999, 2002a,b).

In contrast to the "frozen state" of the mt DNA, the evolution of the nuclear genome in the Metazoan phylum is very fast and follows different pathways in the various organisms (Saccone et al., 2002b; Gissi et al., 2000; Saccone and Pesole, 2003; D'Errico et al., 2004).

The nuclear genome, whose variable size is larger in vertebrates than in invertebrates, is highly redundant due to events of gene duplication or retrotranscription and also to segmental or complete duplication of the whole genome. Redundancy of genetic material is indeed one of the most peculiar features of all genomes, generally increasing with the size and complexity of Metazoan genomes. Both coding and non-coding regions are amplified in a different manner thus creating considerable variability between taxa and even between species. For the coding region, genes having a certain degree of similarity, derived by the above-mentioned duplication events, called paralogs, are grouped in families whose sizes, that is the number of gene members per family, may be different in the various classes (Saccone et al., 2003; Raes and Van De Peer, 2003).

In order to reach a deeper insight into the integrated evolution of the two genomes, nuclear and mitochondrial, in the eukaryotic cell, we have now focused our attention on the genes for mitochondrial products and have started a study of the evolution of nuclearly encoded gene families for mitochondrion in Metazoa (Saccone et al., 2003; Santamaria et al., 2004; Lanave et al., 2004). In the context of this study, our group has developed two specialized databases dedicated to nuclear genes and relevant products (transcripts and amino acid sequences) targeted to mitochondrion, MitoNuc (Attimonelli et al., 2002) and MitoDrome (Sardiello et al., 2003).

In this paper we report the results obtained on the "OXPHOS" genes from several Metazoan species with particular reference to species whose genomes have been completely sequenced.

2. Material and methods

2.1. Grouping OXPHOS genes in ENSEMBL families

SwissProt accession numbers of 78 human OXPHOS proteins were recovered by the MitoDrome database (http://www2.ba.itb.cnr.it/mitodrome) and associated to inter-specific protein families through ENSMART, the annotation tool of the ENSEMBL database (http://www.ensembl.org/Multi/martview). PERL scripts were written to link each ENSEMBL protein family ID to the corresponding genes in all the considered species and to manage data. The adopted ENSEMBL releases are human 26.34.1, chimp 26.1.1, rat 26.3d.1, mouse 26.33a.1, fugu 25.2c.1, zebrafish 26.3.1, drosophila 26.3b.1, mosquito 26.2b.1. The ENSEMBL database implements the Markov cluster algorithm (TRIBE-MCL) for detection of protein families on a large scale (Enright et al., 2002). This method detects and categorizes protein families in the entire genome and among different genomes and relies on a sequence similarity measure obtained through an "all versus all" BLASTp calculation

in the complete genomic set of protein sequences, to consider all relationships in the similarity space at the same time. The same method was adopted to associate the corresponding ENSEMBL family to all predicted genes for each species.

A one-side *t*-Student test on the difference of means with unknown variance was performed, assuming that both OXPHOS families and nuclear gene families represent large (more than 30 elements) and independent samples. A p value lower than 0.0001 allows to reject the null hypothesis that the average OXPHOS family size is greater or equal to the average size of nuclear gene families.

2.2. OXPHOS gene families in Tunicata genomes

A partially manual approach was applied to characterize OXPHOS gene families in two different species of sea squirts (*Ciona intestinalis* and *Ciona savignyi*). The human protein sequences, corresponding to SwissProt IDs reported in Table 1, have been used to scan both *C. intestinalis* genome (JBC, release ver. 1.0, http://genome.jgi-psf.org/ciona4/ciona4.home.html) and *C. savignyi* genome (broad institute, release 1, April 25, 2003 at http://www.broad.mit.edu/annotation/ciona/background.html) by TBLASTN (cutoff threshold of 1, matrix blosum62, Q equal 9 and R equal 2).

3. Results and discussion

3.1. Genome-wide analysis of OXPHOS families

Oxidative phosphorylation is one of the most ancient and conserved functions of all respiring organisms and the components of the respiratory chain complexes are correlated by sub-cellular function and location. Furthermore, respiratory chain complexes are composed of both nuclearly and mitochondrially encoded subunits, with the only exception of complex II, entirely encoded by nuclear genes in Metazoa. Although mt genome has been deeply studied in hundreds of organisms and several Metazoan nuclear genomes are now available and almost entirely annotated, we are just beginning to understand the mechanisms that coordinate these two genomes (Kelly and Scarpulla, 2004). Thus, the analysis of OXPHOS genes may lead to better understand the evolutionary strategies adopted both to preserve OXPHOS function during evolution and to modulate the integrated action of nuclear and mt genomes.

To investigate OXPHOS gene families in Metazoa we used a genome-wide approach in order to produce standardized results that let us compare the size of OXPHOS families, in both an intra- and inter-genomic context. Furthermore, to represent a general evolutionary scenario, species were selected from four different classes (insects, tunicates, fishes, and mammals) and at least two different species were chosen from each class to discriminate between species and class specific features.

The starting set of our analysis consisted of 78 OXPHOS human proteins annotated in the MITODROME database (Sardiello et al., 2003): 66 subunits of the five mitochondrial respiratory chain complexes and 12 proteins associated to complex assembly and OXPHOS functions (accessory proteins).

Inter-genomic families have been characterized using the ENSMART (Kasprzyk et al., 2004; Birney et al., 2004) retrieval tool that automatically associates each protein present in a list to the corresponding family through an “all versus all” intra-and inter-genomes comparison. Although this approach cannot discriminate between genes and pseudogenes, only potential genes able to encode a complete or almost complete protein sequence are taken into account and each gene is strictly associated to a unique family. Since tunicate genome annotation is still missing in the ENSEMBL database, the families identified in the two *Ciona* species are the result of a genome-wide manual screening.

Each gene family consists of one or more proteins encoded by orthologous and paralogous genes in all the investigated species. Table 1 reports the gene family size (number of genes per family) for all characterized OXPHOS gene families in ten species: *Drosophila melanogaster* and *Anopheles gambiae* (insects), *C. intestinalis* and *C. savignyi* (tunicates), *Danio rerio* and *Fugu rubripes* (fishes), *Mus musculus*, *Rattus norvegicus*, *Pan troglodytes* and *Homo sapiens* (mammals).

To investigate the behaviour of OXPHOS genes during evolution, we consider both the ‘OXPHOS families’, that are the families of genes belonging to OXPHOS complexes, and the ‘accessory OXPHOS families’ that are the families formed by genes encoding accessory proteins. The ‘OXPHOS families’ were further divided into complex-specific groups. Table 2 reports the number of conserved subunits with respect to the human set and the family average size, calculated as the ratio between the total number of genes and the number of subunits.

3.2. Evolution of OXPHOS families

The number of conserved subunits of the ‘OXPHOS families’ is 55 for insects, variable from 49 to 62 in tunicates and fishes and almost entirely preserved in mammals (64 in rat, 65 in mouse and 66 in chimp).

The average size of ‘OXPHOS families’ increases passing from insects (1.36 in drosophila and 1.33 in anopheles) to vertebrates (1.55 in zebrafish, 1.45 in fugu, 1.48 in mouse, 1.84 in rat, 1.65 in chimp and 1.61 in man) and the same behaviour is observed for the ‘accessory OXPHOS families’.

These results are consistent with the increase in genome size from insects to vertebrates. Intriguingly, none of the two tunicate genomes presents duplicates of OXPHOS genes. Tunicates are Urochordata, non-vertebrate deuterostomes with some of the smallest genomes (about 180 Mb), and

Table 2
Number of conserved OXPHOS subunits and corresponding family average size

Families	Insects				Tunicates				Fishes				Mammals							
	Ag		Dm		Ci		Cs		Dr		Fr		Mm		Rn		Pt		Hs	
	sb	av	sb	av	sb	av	sb	av	sb	av	sb	av	sb	av	sb	av	sb	av	sb	av
Complex I	27	1.28	27	1.18	31	1.00	26	1.00	24	1.42	29	1.14	32	1.13	31	1.61	32	1.41	32	1.31
Complex II	4	1.00	4	1.75	3	1.00	3	1.00	2	1.50	3	1.25	4	1.00	4	1.25	4	2.50	4	2.25
Complex III	8	1.50	8	1.38	7	1.00	7	1.00	6	1.22	9	1.44	9	1.88	9	2.00	9	1.50	9	163
Complex IV	6	1.17	6	1.50	8	1.00	5	1.00	6	1.50	8	1.88	8	2.50	8	2.70	8	1.88	8	2.00
Complex V	10	1.60	10	1.30	13	1.00	13	1.00	11	1.45	13	1.46	13	1.54	13	2.08	13	2.00	13	1.92
OXPHOS	55	1.36	55	1.33	62	1.00	55	1.00	49	1.55	62	1.45	65	1.48	64	1.84	66	1.65	66	1.61
Accessory OXPHOS	12	1.17	12	1.17	12	1.00	12	1.00	9	1.22	11	1.36	11	1.55	12	1.58	11	1.55	12	1.75
Nuclear genes	–	1.60	–	1.43	–	1.75*	–	–	–	2.98	–	2.24	–	2.18	–	2.22	–	1.80	–	1.89

Light columns report the number of conserved subunits (sb) in each complex (I, II, III, IV, V), in all complexes (OXPHOS) and in the accessory proteins (accessory OXPHOS). Dark columns report the average size (av) of the above-mentioned families and nuclear gene families (nuclear genes).
*Data according to Poustka et al. (2003).

they represent a key reference group between invertebrates and vertebrates. As a matter of fact, Urochordata arose about 800 Myr ago, just before one or two genome-doubling events early at the origin of vertebrates (Vandepoele et al., 2004) and underwent numerous phases of independent lineage-specific gene duplication and loss (Holland and Gibson-Brown, 2003). The complete absence of OXPHOS duplicated genes in tunicates represents an exception in the gene family expansion trend from insects to vertebrates. This implies that *Ciona* genome should have lost duplicated genes, if present, after the separation from the common ancestor, according to the general observation that considerable DNA loss in the tunicate lineage has occurred and a number of genes, e.g. Hox genes, which cluster in most other bilaterians, are uncoupled in *Ciona*. (Holland and Gibson-Brown, 2003). On the other hand, the observed paralogs in insects and vertebrates are probably due to subsequent lineage-specific duplication events.

3.3. OXPHOS families compared to other nuclear families

In order to compare the expansion grade of OXPHOS gene families respect to the global trend of nuclear gene families, we computed the average number of genes per family in the whole nuclear genome, dividing the total number of nuclear genes by the number of associated gene families (see Table 2). It results that the average size of both 'OXPHOS families' and 'accessory OXPHOS families' is smaller than the average size of nuclear gene families in all the considered species (p value <0.0001 in a two-sample t-test between 'OXPHOS families' and nuclear families). On the whole, this suggests that OXPHOS genes are less likely to form duplicates or to preserve them than nuclear genes both in invertebrates and vertebrates.

The low number of OXPHOS paralogs is particularly evident in fish genomes although, as it has been recently demonstrated, this lineage underwent a specific genome-wide duplication (Vandepoele et al., 2004). This event supports and justifies the greater average size of the nuclear gene families in fishes compared to other species (2.8 in zebrafish and 2.24 in fugu), but it has not influenced the size of OXPHOS gene families, for which further duplications should have been selectively lost.

These data are perfectly congruent and extend to other species the observations reported by us on OXPHOS genes in three species of insects. In a recent paper (Tripoli et al., 2005) a detailed analysis has been carried out on the genomes of *D. melanogaster*, *Drosophila pseudobscura* and *Anopheles gambiae* to characterize OXPHOS orthologous and paralogous genes through several criteria: conservation of amino acid sequence, intron/exon structure, introns length and microsyntenic gene order.

This analysis demonstrates that putative orthologs share the identical intron/exon structure in *D. melanogaster* and *D. pseudobscura* in all OXPHOS genes, while the structure is slightly different in *A. gambiae* where exons gain or loss is sometimes observed (see example in Fig. 1). Moreover, the protein sequence identity/similarity values, calculated in OXPHOS genes in these three insects (see example in Table 3), are significantly higher than the average sequence identity of about 56%, reported by Zdobnov et al., 2002 in a comparative genome-wide analysis between orthologous genes of *D. melanogaster* and *A. gambiae*, and the number of duplicates is smaller than expected.

Since we demonstrate that also in vertebrates these genes are not prone to preserve duplicates, we can suppose that the 'balance hypothesis' (Veitia, 2002) is in force in both vertebrates and invertebrates. In agreement with this hypothesis, since protein–protein complex assembly is dosage sensitive, duplicated genes that encode subunits of multi-protein complexes are unlikely to become fixed in the population. Indeed, the imbalance in the concentration of subcomponents should negatively influence their correct assembly and reduce the fitness of organisms. Furthermore, we demonstrate that this behaviour is common also in accessory OXPHOS proteins (see Table 2), thus it seems

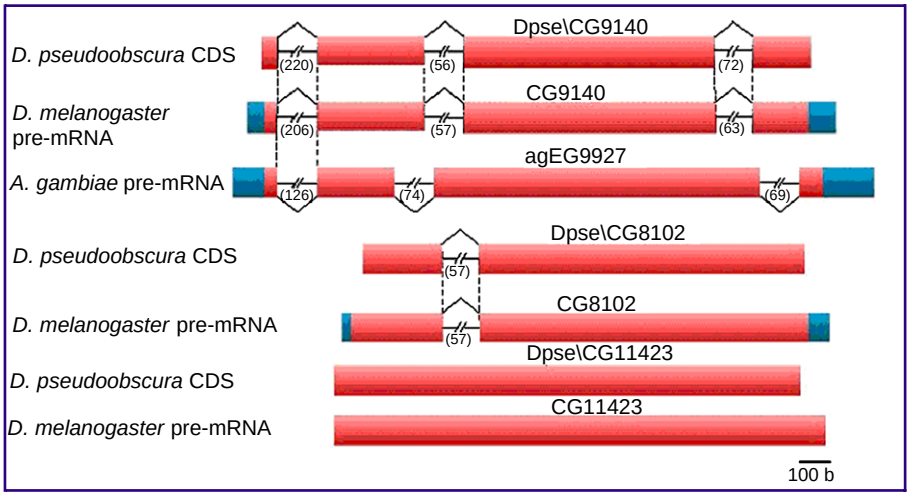

Fig. 1. Comparative analysis of the 51 kDa subunit of NADH–ubiquinone oxidoreductase complex. Structural comparison of intron/exon structure of the gene copies of the 51 kDa subunit of NADH–ubiquinone oxidoreductase complex in *D. melanogaster* (CG9140, CG8102, CG11423), *D. pseudoobscura* (Dpse\CG9140, Dpse\CG8102, Dpse\CG11423) and *A. gambiae* (agEG9927). Homologous coding exons are represented by light boxes, dark boxes are used for UTRs. Length of introns interrupting the coding sequences is reported in round brackets.

that 'balance hypothesis' could constrain also proteins involved in assembly and function of OXPHOS complexes.

Studies on insects showed that the expression of OXPHOS duplicated genes is very low and strongly testis-biased in contrast to the soma-biased expression of the parent genes. These results suggest that, at least in insects, duplicated genes survive only if their expression level is much lower than that of the "founder gene" and the expression is segregated in a specific tissue. We will need further analysis to investigate these mechanisms of regulation in other species and define if both "subfunctionalization" and tissue specificity of duplicated genes (Lynch and Force, 2000) have a conserved role during evolution.

3.4. OXPHOS complex IV and cytochrome c

Although OXPHOS gene families seem smaller respect to the general trend of nuclear gene families, some exceptions are observed. In particular, in most vertebrates the average size of complex IV (cytochrome *c* oxidase) families is larger than that of other complex-specific families and, in mammals, it is also larger than the average size of nuclear gene families. It is already known that some isoforms of specific nuclear cytochrome *c* oxidase subunits exist in a variety of eukaryotic organisms and that this number is variable even from mammal to mammal (Kadenbach and Reinmann, 1998; Bonne et al., 1993). Moreover, it was demonstrated that all the catalytic functions of complex IV are performed by mitochondrially encoded subunits (Moody, 1996). These observations lead us to suppose that OXPHOS functions associated to nuclearly encoded subunits are required to modulate the holoenzyme activity in a tissue-specific or a developmentally regulated manner (Burke and Poyton, 1998) and that the high level of paralogy for these complex subunits could be part of this complicated plot.

Intriguingly, also cytochrome *c* represents a wide gene family with at least four members in each mammal but only one or two members in insects and tunicates (Table 1). The selective amplification of cytochrome *c* gene copies in such complex organisms as mammals could be correlated to the involvement in other cellular processes, for example apoptosis, in which cytochrome *c* participates in mammals but not in insects (Dorstyn et al., 2004).

Table 3
Pair-wise alignment of 51 kDa subunit of NADH–ubiquinone oxidoreductase complex

Gene name	*D. melanogaster*		*D. pseudoobscura*			*A. gambiae*
	CG8102	CG11423	Dpse\CG9140	Dpse\CG8102	Dpse\CG11423	AGeg9927
CG9140	45/59	64/71	**95/97**	49/64	68/76	**84/90**
CG8102		40/50	46/59	68/81	42/55	45/58
CG11423			65/75	43/55	77/84	64/72
Dpse\CG9140				48/64	65/73	**84/90**
Dpse\CG8102					43/55	48/63
Dpse\CG11423						64/71

Pair-wise alignments are relevant to unprocessed predicted protein sequences.
Columns report identity/similarity percentage ratio values of pair-wise alignments. More significant values are reported in bold.

3.5. Evolution of other nuclear genes

In previous papers (Saccone et al., 2003; Santamaria et al., 2004) we have investigated the evolution of other nuclear proteins, namely the porin (VDAC) and the adenine nucleotide transporter (ANT) located in, respectively, the outer and inner membrane of the mitochondrion. We have found that both invertebrates and vertebrates possess duplicated genes. The VDAC gene redundancy found in invertebrates and possibly in some fishes may indicate a tendency to duplicate the genetic material, rather than a real need for function innovation. It is conceivable that in the various lineages, paralogous genes have followed various fates, some have been conserved, some have been lost, and others have been amplified by further duplications. Another relevant result is that even closely related species can have a different number of genes. In humans for example, there are three ANT isoforms, each differently regulated with a tissue-specific expression pattern. However in some mammalian species, not all three forms are present: one form, ANT3, is absent in rodents and another, ANT2, is missing in artiodactyls. The future challenge will be to discover the functions kept or gained by the various isoforms during evolution.

3.6. Conclusions

The results reported here clearly demonstrate that both in insects and vertebrates the genes for OXPHOS complexes and for accessory OXPHOS proteins are modestly duplicated.

The average size of OXPHOS families increases going from insects to mammals, but remains smaller than the size of other nuclear gene families in all considered species.

This trend is in line with the ‘balance hypothesis’, put forward for insects, according to which duplicated genes coding for multi-protein complexes are negatively selected, since their expression may influence the complex formation and thus the fitness of organisms. The low expression level of the duplicated genes in insects is consistent with this hypothesis.

More detailed analyses on both genes families and on their expression in other Metazoa, vertebrates and invertebrates, remain to be carried out for drawing general conclusions.

Acknowledgments

This work has been supported by “Ministero dell’Istruzione, dell’Università e della Ricerca”-Italy. Projects: Cluster C03-Progetto n. 2-L. 488/92 :“Studio di Geni di interesse Biomedico ed Agroalimentare”; Progetto Strategico Genetica molecolare, legge 449/97; “Centro di Eccellenza in Genomica Comparata” (CEGBA).

References

Attimonelli, M., et al., 2002. MitoNuc: a database of nuclear genes coding for mitochondrial proteins. Nucleic Acids Res. 30, 172–173 (Update 2002).

Birney, E., et al., 2004. ENSEMBL 2004. Nucleic Acids Res. 32, D468–D470 (Database issue).

Bonne, G., Seibel, P., Possekel, S., Marsac, C., Kadenbach, B., 1993. Expression of human cytochrome *c* oxidase subunits during fetal development. Eur. J. Biochem. 217, 1099–1107.

Burke, P.V., Poyton, R.O., 1998. Structure/function of oxygen-regulated isoforms in cytochrome *c* oxidase. J. Exp. Biol. 201, 1163–1175.

D’Errico, I., Gadaleta, G., Saccone, C., 2004. Pseudogenes in Metazoa: origin and features. Brief. Funct. Genomic Proteomic 3, 157–167 (Review).

Dorstyn, L., Mills, K., Lazebnik, Y., Kumar, S., 2004. The two cytochrome *c* species, DC3 and DC4, are not required for caspase activation and apoptosis in Drosophila cells. J. Cell Biol. 167, 405–410.

Enright, A.J., Van Dongen, S., Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584.

Gissi, C., Reyes, A., Pesole, G., Saccone, C., 2000. Lineage-specific evolutionary rate in mammalian mtDNA. Mol. Biol. Evol. 17, 1022–1031.

Holland, L.Z., Gibson-Brown, J.J., 2003. The *Ciona intestinalis* genome: when the constraints are off. Bioessays 25, 529–532.

Kadenbach, B., Reinmann, A., 1998. Cytochrome *c* oxidase: tissue specific expression of isoformes and regulation of activity. In: Ernster, L. (Ed.), Molecular Mechanism of Bioenergetics, pp. 241–243.

Kasprzyk, A., et al., 2004. EnsMart: a generic system for fast and flexible access to biological data. Genome Res. 14, 160–169.

Kelly, D.P., Scarpulla, R.C., 2004. Transcriptional regulatory circuits controlling mitochondrial biogenesis and function. Genes Dev. 18, 357–368.

Lanave, C., Santamaria, M., Saccone, C., 2004. Comparative genomics: the evolutionary history of the Bcl-2 family. Gene 333, 71–79.

Lynch, M., Force, A., 2000. The probability of duplicate gene preservation by subfunctionalization. Genetics 154, 459–473.

Margulis, L., 1970. Origin of Eukaryotic Cells. Yale University Press, New Haven, Conn.

Martin, W., Muller, M., 1998. The hydrogen hypothesis for the first eukaryote. Nature 392, 37–41.

Moody, A.J., 1996. ‘As prepared’ formes of fully oxidized haem/Cu terminal oxidases. Biochim. Biophys. Acta 1253, 6–20.

Poustka, A.J., et al., 2003. Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters. Genome Res. 13, 2736–2746.

Raes, J., Van De Peer, Y., 2003. Gene duplication, the evolution of novel gene functions, and detecting functional divergence of duplicates in silico. Appl. Bioinformatics. 2, 91–101 (Review).

Saccone, C., 1994. The evolution of mitochondrial DNA. Opin. Genet. Dev. 4, 875–881.

Saccone, C., Pesole, G., 2003. Handbook of Comparative Genomics: Principles and Methodology. Wiley, Hoboken, NY, USA.

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G., Reyes, A., 1999. Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238, 195–209 (Review).

Saccone, C., Gissi, C., Reyes, A., La rizza, A., Sbisa, E., Pesole, G., 2002a. Mitochondrial DNA in Metazoa: degree of freedom in a frozen event. Gene 286, 3–12 (Review).

Saccone, C., et al., 2002b. Molecular strategies in Metazoan genomic evolution. Gene 300, 195–201.

Saccone, C., Caggese, C., D’Erchia, A.M., Lanave, C., Oliva, M., Pesole, G., 2003. Molecular clock and gene function. J. Mol. Evol. 57 (Suppl 1), S277–S285.

Santamaria, M., Lanave, C., Saccone, C., 2004. The evolution of the adenine nucleotide translocase family. Gene 333, 51–59.

Sardiello, M., Licciulli, F., Catalano, D., Attimonelli, M., Caggese, C., 2003. MitoDrome: a database of *Drosophila melanogaster* nuclear genes encoding proteins targeted to the mitochondrion. Nucleic Acids Res. 31, 322–324.

Tripoli, G., D'Elia, D., Barsanti, P., Caggese, C., 2005. Comparison of the oxidative phosphorylation (OXPHOS) nuclear genes in the genomes of *Drosophila melanogaster*, *Drosophila pseudoobscura* and *Anopheles gambiae*. Genome Biol. 6 (2), R11.

Vandepoele, K., De Vos, W., Taylor, J.S., Meyer, A., Van de Peer, Y., 2004. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. Proc. Natl. Acad. Sci. 101, 1638–1643.

Veitia, R.A., 2002. Exploring the etiology of haploinsufficiency. Bioessays 24, 175–184.

Zdobnov, E.M., et al., 2002. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. Science 298, 149–159.