



Evolution of residues from the cytochrome c oxidase complex engaged in intermolecular contacts

Journal:	<i>Molecular Biology and Evolution</i>
Manuscript ID:	MBE-13-0185.R1
Manuscript Type:	Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Valverde, Héctor; Universidad de Málaga, Biología Molecular y Bioquímica Ruíz-Camacho, Manuel; Universidad de Málaga, Estadística e Investigación Operativa López, Francisco; Universidad de Málaga, Estadística e Investigación Operativa Morilla, Ian; Universidad de Málaga, Biología Molecular y Bioquímica Aledo, Juan; Universidad de Málaga, Biología Molecular y Bioquímica
Key Words:	Protein-protein interaction, mtDNA, Coevolution, COX, Natural selection, Protein evolution

SCHOLARONE™
Manuscripts

Research Article (Discoveries)

**Evolution of residues from the cytochrome c oxidase complex
engaged in intermolecular contacts**

Héctor Valverde^a, Manuel Ruíz-Camacho^b, Ian Morilla^a, Francisco Demetrio
López^b, Juan Carlos Aledo^{a*}

^aDepartamento de Biología Molecular y Bioquímica. Facultad de Ciencias. Universidad de
Málaga, 29071 Málaga, Spain

^bDepartamento de Estadística e Investigación Operativa. Facultad de Ciencias. Universidad de
Málaga, 29071 Málaga, Spain.

*Corresponding author:

Email: caledo@uma.es

Phone: +34952137129

Keywords: *Coevolution, COX, evolvability, mtDNA, natural selection, protein evolution, protein-protein interaction.*

Abstract

Respiratory complexes are encoded by two genomes (mtDNA and nDNA). Although the importance of intergenomic coadaptation is acknowledged, the forces and constraints shaping such coevolution are largely unknown. Previous works using cytochrome c oxidase (COX) as a model enzyme, have led to the so-called ‘optimizing interaction’ hypothesis. According to this view, mtDNA-encoded residues close to nDNA-encoded residues evolve faster than the rest of positions, favouring the optimization of protein-protein interfaces. Herein, using evolutionary data in combination with structural information of COX, we show that failing to discern the effects of interaction from other structural effects, can lead to deceptive conclusions such as the ‘optimizing hypothesis’. Once spurious factors have been accounted for, data analysis shows that mtDNA-encoded residues engaged in contacts are, in general, more constrained than their noncontact counterpart. Nevertheless, noncontact residues from the surface of COX 1 subunit are a remarkable exception, being subjected to an exceptionally high purifying selection. We also report that mtDNA-encoded residues involved in contacts with other mtDNA-encoded subunits, are more constrained than mtDNA-encoded residues interacting with nDNA-encoded polypeptides. This differential behaviour cannot be explained on the basis of thermodynamic stability, since interactions between mtDNA-encoded subunits contribute more weakly to the complex stability than those interactions between subunits encoded by different genomes. Finally, among nDNA-encoded subunits, contact residues are more conserved than noncontact residues, being COX 5A an exception.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Introduction

Although mitochondria are involved in many aspects of cell function, including proliferation (Aledo 2004), apoptosis (Suen et al. 2008) and aging (Aledo et al. 2010), their central role is related to energy transduction in oxidative phosphorylation (OXPHOS). The mitochondrial proteins responsible for the OXPHOS are encoded by two genomes. In mammals, the mitochondrial genome (mtDNA) encodes for 13 polypeptides that interact with a large number of nuclear-encoded (nDNA-encoded) polypeptides to form the functional complexes I, III, IV and V of the OXPHOS system. Given that each mtDNA gene product must interact with proteins encoded by the nuclear genome to carry out its functions, coevolution between mtDNA and nDNA leading to intergenomic coadaptation is expected.

As a consequence of their critical function, mutations altering the structure of these oligomeric complexes must face the close scrutiny of natural selection. In this way, natural selection will favour evolutionary coadaptation of interacting proteins, either to improve physiological functions (Schmidt et al. 2005), or just to maintain the fitness through compensatory changes after a slightly deleterious mutation has been fixed by genetic drift (Osada and Akashi 2011). Whatever the driving force may be, there are numerous examples of the biological importance of intergenomic coadaptation. In this sense, xenomitochondrial cybrid cells constructed using nDNA from one species and mtDNA from a close species were viable and had a functional OXPHOS, whereas more divergent species failed to produce functional OXPHOS complexes (Kenyon and Moraes 1997; Barrientos et al. 2000; McKenzie 2003). These results underline the importance of cytonuclear coevolution. Similar conclusions were derived from studies where repeated backcrossing of genetically isolated populations of the copepod *Tigriopus californicus*, allowed to place the maternally inherited mtDNA genome of one population with the paternal nDNA of another. These interpopulation hibrids exhibited a

defective OXPHOS system (Edmands and Burton 1999). A third line of evidence suggesting that OXPHOS proteins have coevolved to function optimally, comes from epistatic studies where mutations in mtDNA genes that are pathogenic in humans have been observed in naturally occurring genomes from nonhuman mammals (de Magalhães 2005; Azevedo et al. 2009).

While the above studies emphasize the relevance of coevolution between interacting proteins, they do not provide much insight on the forces and constraints shaping such coevolution. In an attempt to explore the evolutionary dynamics of these protein-protein interactions, Schmidt and co-workers used cytochrome c oxidase (COX) as a model of OXPHOS holoenzyme. These authors analysed the rate of nonsynonymous substitutions within a set formed by mtDNA-encoded residues in physical proximity to nDNA-encoded amino acids, and compared it with that computed for the rest of mtDNA-encoded residues, which are not in contact with nDNA-encoded polypeptide chains. They concluded that mtDNA-encoded residues in close contact with amino acids being encoded by the nucleus, evolve faster than the rest of mtDNA-encoded residues (Schmidt et al. 2001). This result was interpreted as being due to many different amino acid replacements among the close contact residues being required to optimize this protein's interaction with other proteins. Actually, the authors referred to such a state as an "optimizing interaction". In contrast, when COX residues encoded by nDNA were segregated on the basis of proximity to mtDNA-encoded residues, and the rates of nonsynonymous substitution were analysed, the conclusion reached was the opposite. That is, those nDNA-encoded residues in contact with mtDNA-encoded amino acids, evolve more slowly than the rest of nDNA-encoded residues (Schmidt et al. 2001).

These striking results have been often cited as an example of the differential forces driving mtDNA and nDNA evolution (Willett 2003; Das et al. 2004; Castellana et al. 2011). While the constrained evolution of nDNA-encoded interacting residues is in line with the

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

97 prevailing view that protein evolution is generally conservative and constraining interactions
98 are typical, the observation that mtDNA-encoded interacting residues evolve at much higher
99 rates than non interacting amino acids, if confirmed as a bona fide observation, deserved a
100 sound explanation. Herein, we have revisited the “optimizing interaction” hypothesis. Using a
101 comprehensive number of mammalian taxa and extended statistic analyses (Fig. 1), we have
102 found that interacting mtDNA-encoded residues, alike interacting nDNA-encoded residues, are
103 subjected to higher constraints than their corresponding non-interacting counterparts. We also
104 provide the keys to understand why previous studies failed to reach similar conclusions. In
105 addition, we show an intriguing difference in the evolutionary rate of mtDNA-encoded residues
106 depending on whether they contact with nDNA-encoded subunits or with other mtDNA-
107 encoded subunits.

108 **Results and discussion**

109 *Discerning the effects of residue interaction on the evolutionary rate from other structural*
110 *effects.*

111 The optimizing interaction hypothesis arose from the observation that mtDNA-encoded
112 residues from COX in close contact with nDNA-encoded amino acids, showed a higher rate of
113 nonsynonymous substitutions than the rest of mtDNA-encoded residues (Schmidt et al. 2001).
114 Since these rates of nonsynonymous substitutions were originally calculated using orthologous
115 sequences from only 26 mammalian species, we wanted to start by re-assessing this issue using
116 a more comprehensive collection of sequences from 371 mammalian taxa. In this way, using
117 our extensive alignment and the methodology described by Schmidt et al., the set formed by the
118 aligned codons from the three mtDNA-encoded COX subunits (ABC), was split into two
119 subsets: $ABC_{Mt-nu_Contact}$ and its complement, $(ABC_{Mt-nu_Contact})^c$. The former encompassed those
120 triplets encoding for residues close to nDNA-encoded aminoacids in the holoenzyme, while the

latter set contained all the codons from chain A, B and C, which encoded amino acids that are not in contact with nDNA-encoded residues (Fig. 2). The calculated interaction ratio, $R_{ABC_{Mt-nu_Contact}}$, was $1.805 \pm 4 \cdot 10^{-4}$, significantly greater than 1. Although this result is in line with the data reported by Schmidt and coworkers, such an observation by itself is, in our opinion, insufficient to support the conclusion that contact residues are subjected to a positive selection, as suggested in previous works (Schmidt et al. 2001).

In this sense, a number of considerations need to be addressed before any conclusion can be reached. For instance, the set formed by mtDNA-encoded residues that are not in contact with nDNA-encoded amino acids, used as reference to compute the interaction ratio, represents a heterogeneous collection of residues. Thus, while the amino acids belonging to the $ABC_{Mt-nu_Contact}$ group are mainly located at the protein surface, a significant part of the noncontact residues are buried into the protein structure (Fig. 2). This observation is relevant because we have recently reported that, for mtDNA-encoded proteins, alike proteins from other genetic origins, buried residues are most likely to remain conserved during evolution compared to their solvent exposed counterparts (Aledo et al. 2012). Therefore, if we want to address the effects of residue interaction on the evolutionary rate, and discern them from other structural effects such as the solvent exposure, the noncontact set used as reference should be restricted to avoid buried residues.

Although the above mentioned restriction is necessary, yet it is not sufficient to build a suitable reference set. Indeed, the resulting set of such a restriction still contains a group of residues that may bias the results and mislead the conclusions. We are referring to the collection formed by those amino acids implicated in protein-protein interactions involving only contacts between mtDNA-encoded residues (see $ABC_{Mt-mt_Contact}$ in Figure 2). Since the evolvability of this category of residues has not been previously characterized, we excluded them from the reference set, which finally was formed only by those mtDNA-encoded residues

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

146 exposed at the protein surface that are not involved in any sort of intersubunit contacts. This
147 reference set is referred to as $ABC_{Exposed_Noncontact}$ (Fig. 2).

148 Once the partition of the initial data set had been carried out as described above and
149 illustrated in Figure 2, we next computed diverse evolutionary variables to characterize the
150 relative evolvability of the following three sets: $ABC_{Mt-nu_Contact}$, $ABC_{Mt-mt_Contact}$ and
151 $ABC_{Exposed_Noncontact}$, all of them containing only codons for amino acids exposed at the protein
152 surface. The results of such analyses are described next.

153 ***Differential behaviour between Mt-mt and Mt-nu interactions.***

154 To assess the relative evolvability of each category of residue we used different
155 approaches. One of these, consisted in calculating the so-called relative evolutionary rate
156 (RER). Briefly, for each mtDNA-encoded subunit, the RER of each position was computed
157 using MEGA 5 (Tamura et al. 2011), which implements a maximum likelihood method to
158 model evolutionary rate differences among sites. These rates were scaled such that the
159 average evolutionary rate across all sites is 1. In this way, sites showing a rate lower than 1
160 are evolving slower than average, while those with a rate higher than 1 are evolving faster
161 than average. As it can be observed in Table 1, among all exposed residues, those involved in
162 interactions between different mtDNA-encoded chains were the most constrained, showing
163 RER values well below the unit, regardless the mtDNA-encoded subunit being considered
164 (Table 1). This observation indicates a much stronger purifying selection among Mt-mt
165 Contact residues than among Mt-nu Contact amino acids. To substantiate this conclusion, a
166 different approach was followed. For each of the three subsets of exposed residues (Exposed
167 Noncontact, Mt-nu Contact and Mt-mt Contact) as well as for the Buried Noncontact subset,
168 the number of nonsynonymous substitutions per nonsynonymous site, dN, was calculated
169 from pairwise comparisons. The sum of these nonsynonymous sequence divergences for all

the pairwise comparisons was computed and denoted as ΣdN . In line with the results based on RER, the Mt-mt Contact set exhibited the lower ΣdN values among all the exposed residues (Table 1). Since dN is informative of the combined effect of mutation and selection, we used ΣdN as a proxy of the evolvability of the corresponding subset of residues being analysed. However, ΣdN averages over all the analysed mammalian species and, therefore, it provides little information on whether the evolutionary dynamics described above, has a broad phylogenetic distribution and is present in most mammalian lineages. To address this issue, phylogenetic trees were used to apportion the nonsynonymous substitutions for the different residue subsets. As it can be deduced from Fig. 3, all lineages without exception showed much higher dN when the analysed subset was mtDNA-encoded residues contacting with nDNA-encoded subunits, with respect to those contacting with other mtDNA-encoded polypeptides. In most cases, the difference in the dN values computed for these two groups of residues was of one order of magnitude. (See also Supplementary Material-1). All together, these results suggest that those mtDNA-encoded residues in contact with mtDNA-encoded residues from a different chain, are subjected to stronger constraints than those involved in interactions with nDNA-encoded subunits. In addition, this behaviour seems to have a broad phylogenetic distribution and is valid for most mammalian lineages (Fig. 3). To the best of our knowledge, this is the first quantitative study supporting such a conclusion, which may be rationalized as follows. In mammals, the three mtDNA-encoded subunits form the catalytic core of the enzyme, while the 10 nDNA-encoded subunits act as a regulatory shield surrounding the core (Soto et al. 2012). Prokaryote forms of COX boil down to the catalytic core, which seems to have an ancient origin (Castresana et al. 1994). In contrast, eukaryotes possess nuclear genes for additional subunits, the number of which generally increases with the organismal complexity. Although, neither the origin nor the specific function of these additional non-catalytic subunits is completely understood, there is little doubt that they are

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

significantly younger than the proteins conforming the catalytic core (Castresana et al. 1994; Das et al. 2004; Little et al. 2010). On the other hand, it is well known that young proteins tend to experience weaker purifying selection and evolve more quickly than old proteins (Alba and Catresana 2005; Vishnoi et al. 2010). Herein, we propose that what is true for proteins may also be true for interactions. In other words, it is reasonable to assume that within a given protein with a fixed antiquity, those residues involved in “old interactions” evolve more slowly than those other residues implicated in “young interactions”. Describing how selection pressure acts at the interfaces of protein-protein complexes is a fundamental issue with high interest for the structural prediction of macromolecular assemblies. Therefore, although it is out of the scope of the current research, in the future it would be interesting to assess whether the relationship between the strength of selection and the age of the interaction, described herein for mtDNA-encoded COX subunits, also applies to other proteins. To encourage this task, we have released the program that implement part of the workflow sketched in Fig. 1. This program has been submitted into the public repository CPAN under an open source license, which allows any other researcher to modify the software in order to include other structural, evolutionary and statistical analyses that may help to gain insights into the molecular evolution of any particular protein complex of interest. Documentation and guides for users and developers are available at the program website <http://mecom.hval.es>.

Thermodynamic stability does not account for the differential behaviour between Mt-mt and Mt-nu interactions.

Once we had established that residues from the Mt-mt Contact group are much more constrained than those belonging to the Mt-nu Contact set, we wondered if such differential pattern might have arisen from a differential contribution of these two categories of contacts to the stability of the complex. To address this issue, we carried out an *in silico* alanine scanning

mutagenesis. The results of such analysis are summarized in Table 2. As expected, changes affecting noncontact residues were among the least destabilizing mutations. On the other hand, from a thermodynamic point of view, residues from the Mt-nu Contact class were much less tolerant to changes than any other kind of residue, including those from the Mt-mt Contact set, which exhibited an intermediate behaviour. Since mutations between residues from the Mt-mt Contact class tend to be less destabilizing than those affecting residues from the Mt-nu Contact group, the higher degree of conservation observed between Mt-mt Contact residues, does not seem to be based on thermodynamic stability, suggesting that functional aspects may be behind the high degree of conservation observed among Mt-mt contact residues.

The exposed noncontact residues from COX 1 are exceptionally conserved.

Regardless of the genetic origin of the contacted residue, interacting mtDNA-encoded residues seem to be more constrained than their non interacting counterparts, as suggested by an interaction ratio below the unit ($0.8 \pm 2 \cdot 10^{-4}$). However, when each individual chain was separately analysed, COX 1 showed a unique behaviour. Thus, while the interaction ratios for COX 2 and COX 3 were significantly lower than 1, COX 1 exhibited a value significantly higher than 1. The departure of COX 1 from the general trend, may be due to either an increased rate of nonsynonymous substitutions between codons within the Contact set, which would favour the “optimizing hypothesis”, or to a reduced rate of nonsynonymous substitutions between codons belonging to the Exposed Noncontact set. To address this issue, we next computed and plotted ΣdN for the Contact and Exposed Noncontact sets of each mtDNA-encoded chain (Fig. 4). From this figure, it becomes evident that i) changes in COX 1 are much more constrained than in any other chain, and ii) this was particularly true within the exposed noncontact group, arguing against the “optimizing hypothesis” even in the case of COX 1.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

From the comparison of ΣdN between subunits, it becomes clear-cut that exposed noncontact residues from COX 1 are particularly constrained (Fig. 4). However, if we compare the inertia to changes of this subset of residues with respect to the rest of COX 1 residues, we may wonder whether exposed noncontact residues from COX 1 are significantly more conserved than the rest of COX 1 residues. To address this question while avoiding assumptions about the underlying distributions, we resorted to a bootstrap approach. Briefly, for each mtDNA-encoded chain, the codons from the multiple sequence alignment were randomly sorted to form a subset of the same size than the original Exposed Noncontact subset of the corresponding chain. Afterwards, ΣdN was computed using this random subset. For each chain, the random resampling was performed 10^4 times to build up empirical distributions, which were used to contrast the ΣdN values computed in the real Exposed Noncontact subsets. As it can be deduced from Fig. 5, those residues belonging to the Exposed Noncontact groups from COX 2 and COX 3 are among the most variable residues (p-values 0.003 and 0.045, respectively). In contrast, exposed noncontact residues from COX 1 were among the most conserved residues (Fig. 5).

In an attempt to get further insight into the particular forces imposing such exceptionally high degree of conservation between exposed noncontact COX 1 residues, we assessed the effect of *in silico* alanine scanning mutagenesis on the stability of the complex. Although the mean $\Delta\Delta G$ for exposed noncontact COX 1 residues (1.64 kJ/mol) was significantly higher than those values from COX 2 (1.20 kJ/mol) and COX 3 (1.16 kJ/mol), p-values 0.015 and 0.008, respectively, the contribution of COX 1 exposed noncontact residues to the whole thermodynamic stability of the holoenzyme can hardly be invoked as a reason for their high degree of conservation, as it is evident from Fig. 6. For each mtDNA-encoded protein, the Contact and Exposed Noncontact subsets were employed to compute their ΣdN values, which were plotted against their corresponding $\Delta\Delta G$ mean values (Fig. 6). We found a

268 significant (p -value = 0.035) negative correlation between these two variables, indicating that
269 those substitutions that tend to be more destabilizing are also more constrained, which is in line
270 with our previous observations that thermodynamic stability plays a relevant role in the
271 evolvability of mtDNA-encoded proteins (Aledo et al. 2012). However, the COX 1 Exposed
272 Noncontact group showed up as an outlier. In this sense, for a $\Delta\Delta G$ of 1.64 kJ/mol (the mean
273 value computed for COX 1 exposed noncontact residues) the expected ΣdN should be higher
274 than twice the observed ΣdN (Fig 6). In other words, whatever the constraining forces may be,
275 they seem to be unrelated to structural stability.

276 Thus, we next explored other alternatives. In this sense, it is widely acknowledged that
277 functional regions of proteins exhibit higher inertia to nonsynonymous changes than those other
278 regions (Glaser et al. 2003). Hence, if there were a higher proportion of functionally important
279 residues in the exposed noncontact region of COX 1 than in its contact counterpart, this would
280 help to explain the extraordinarily low ΣdN computed for the set Exposed Noncontact from
281 COX 1. To test this possibility, we focused on those COX 1 residues that have been described
282 to fulfil important functions such as the binding of heme, Cu and Mg, proton pumping and
283 electron transfer (Michel et al. 1998; Yoshikawa 1998). Figure 6 inset shows that the
284 proportion of these residues that belong to the Exposed Noncontact set override the proportion
285 of those that pertain to the Contact group. Although this observation is qualitatively in line with
286 the very low ΣdN observed for the COX 1 Exposed Noncontact ensemble, it should be noted
287 that the catalogue of COX 1 functional residues we have used, is very limited in size and
288 probably it is far from the complete inventory, which hampers obtaining quantitative statistical
289 support. Nevertheless, the exceptionally high degree of conservation observed between exposed
290 noncontact COX 1 residues, likely captures important constraints that apply to biological
291 functionality yet to be unravelled.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

292 In this regard, we wondered whether this set of COX 1 exposed noncontact residues
293 might be involved in functions related to the formation of mitochondrial supercomplexes.
294 Indeed, during the past decade, significant experimental evidence supports the organization of
295 the mitochondrial respiratory chain into higher order structures known as respirasome
296 (Schägger and Pfeiffer 2000; Acín-Pérez et al. 2008; Winge 2012). Along the same lines, two
297 quasi-simultaneous recent studies have reported the 3D structure of mitochondrial
298 supercomplex I₁III₂IV₁, determined by electron cryo-microscopy at 19-22 Å resolution (Althoff
299 et al. 2011; Dudkina et al. 2011). Fitting of X-ray structures of single complexes I, III₂ and IV
300 with high fidelity, unravels only a few sites where neighbouring complexes come close enough
301 for ion bridges or hydrogen bonds. Three of such sites of potentially strong protein-protein
302 interaction were found between complexes III and IV. However, none of these sites were on
303 COX 1 (Althoff et al. 2011; Dudkina et al. 2011). Nevertheless, beside the supercomplex
304 I₁III₂IV₁, other forms of supercomplexes such as I₁III₂IV₂, I₁III₂IV₃ and I₁III₂IV₄, have been
305 described in bovine heart mitochondria (Schägger and Pfeiffer 2001). Since the three
306 dimensional structures of these less abundant forms of supercomplexes are unknown, the
307 involvement of residues from COX 1 in the assembly/stability of respirasomes, although
308 unlikely, cannot be completely ruled out. In any event, the high degree of conservation between
309 the exposed noncontact residues from COX 1 herein described, is an unexpected and intriguing
310 observation that we are currently unable to explain. Nevertheless, we have noted that when this
311 particular behaviour of COX 1 noncontact residues is not properly accounted, it leads to high
312 interaction ratios that may be misinterpreted as an optimizing high evolvability of Mt-nu
313 contact residues.

314 *Contact residues from nDNA-encoded proteins tend to be more constrained than their*
315 *noncontact counterparts.*

1
2
3 316 A number of works have previously addressed the evolution of nDNA-encoded COX
4
5 317 subunits (Schmidt et al. 2001; Das et al. 2004; Uddin et al. 2008; Osada and Akashi 2011).
6
7 318 However, some of the conclusions drawn from these studies are in apparent conflict. For
8
9 319 instance, Schmidt and coworkers concluded that while contact nDNA-encoded residues are
10
11 320 subjected to a strong purifying selection (Schmidt et al. 2001), Osada and Akashi reached the
12
13 321 conclusion that nDNA-encoded residues exhibit signs of positive selection (Osada and Akashi
14
15 322 2011). These conflicting results may be attributable to methodological differences. Thus, while
16
17 323 the former authors assembled and analysed a data set formed by contact and noncontact
18
19 324 residues without distinguishing between different COX subunits, the latter authors looked for
20
21 325 individual sites under positive selection using the methodology described somewhere else
22
23 326 (Zhang 2005). Herein, we have readdressed the issue employing a meso approach, consisting in
24
25 327 analysing each individual nDNA-encoded chain.
26
27
28
29

30 328 As a first approach, we computed the interaction ratio for each subunit. Contact residues
31
32 329 tend to be more conserved than their noncontact counterpart, as suggested by R_{12} values well
33
34 330 below the unit, with the exception of COX 5A (Fig. 7A). This subunit is much more conserved
35
36 331 than any of the other nuclear-encoded subunits (Fig. 7B). The high level of conservation, which
37
38 332 extends beyond the contact residues, suggests that subunit 5A has been subjected to an
39
40 333 exceptionally strong purifying selection. Supporting this implication of functional constraint is
41
42 334 experimental evidence for ligand interaction that involves subunit COX 5A as a receptor of 3,5-
43
44 335 diiodothyronine, able to abolish the allosteric inhibition of respiration by ATP (Arnold et al.
45
46 336 1998).
47
48
49
50

51 337 In general lines, the results summarized in Fig. 7A are consistent with those obtained by
52
53 338 other authors using the pooled sequences (Schmidt et al. 2001). In analysing each individual
54
55 339 nDNA-encoded polypeptide, we aimed to detect those subunits that might exhibit a particular
56
57 340 evolutionary behaviour. However, such goal faced methodological challenges. Probably, the
58
59
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

most important methodological limitation was the impossibility, because of the small size of the nDNA-encoded COX subunits, of obtaining a statistically relevant sample of residues. Therefore, in the absence of a large enough sample of residues, the herein computed R_{12} values can only give a qualitative indication of the trend, but they lack of statistical significance. To overcome such inconvenience, we resorted to nonparametric analyses.

For a given subunit, each residue was labeled as ‘variant’ or ‘invariant’, depending on its conservation throughout a multisequence alignment (Supplementary Material-2). On the other hand, according to the criterion already described, the same residue was classified as ‘contact’ or ‘exposed noncontact’. In this way, the proportion of variant exposed noncontact residues (p_0) and that of variant contact residues (p_1) were computed, as well as its differences, $d = p_0 - p_1$. According to a null hypothesis where the variability of a residue is independent of its role in intersubunit interaction, these differences should fit to a symmetric distribution with zero median. To test such a hypothesis, a Wilcoxon signed-rank test was carried out, resulting in the rejection of the null hypothesis ($p < 0.007$). Thus, the higher proportion of variable residues among the exposed noncontact positions with respect to the contact positions (see Table from Supplementary Material-3) are unlikely to be due to chance.

Concluding remarks

Interactions of mtDNA- and nDNA-encoded proteins provide unique opportunities to study the evolution of protein-protein interactions and the effects of these interactions on the evolution of their respective genomes. To this respect, mammalian COX complex is well suited for studying the evolution of within- and between-genome interactions because a functional complex requires three mtDNA-encoded and ten nDNA-encoded subunits, accounting for a large number of interacting residues. Previous studies carried out with this

complex, have suggested that mtDNA-encoded residues in close physical contact with nDNA-encoded amino acids may be subjected to positive selection to optimize the structural-functional interaction between subunits from different genetic origins. This conclusion was based on the higher rate of nonsynonymous substitutions observed among mtDNA-encoded residues in contact with nDNA-encoded residues, with respect to the rest of mtDNA-encoded amino acids. However, as we have shown herein, failing to discern the effect of interaction from other confounding factors such as the solvent exposure, can lead to misleading conclusions. Thus, when such corrections were made, data analysis showed that mtDNA-encoded residues engaged in contacts are, in general, more constrained than their noncontact counterparts. Nevertheless, noncontact residues from the surface of COX 1 subunit are a remarkable exception, being subjected to an exceptionally high purifying selection.

Besides providing compelling evidence against the so-called optimizing interaction hypothesis, our approach also allowed to make some interesting findings. For instance, those mtDNA-encoded residues in contact with mtDNA-encoded residues from a different chain, are subjected to stronger constraints than those involved in interactions with nDNA-encoded subunits. This observation cannot be explained on the basis of thermodynamic stability, because interactions between mtDNA-encoded subunits contribute more weakly to the complex stability than those interactions between subunits encoded by different genomes. Therefore, the higher conservation observed among mtDNA-encoded residues involved in within genome interactions, is likely due to functional rather than structural reasons.

Materials and Methods

DNA Sequences

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

A collection of 371 mammalian mitochondrial genomes was obtained from the National Center for Biotechnology Information (NCBI) genome database (www.ncbi.nlm.nih.gov). For the analyses involving COX subunits encoded by nuclear genes, sequences from a variable number of species ranging from 14 to 30 (depending on sequence availability) were acquired from NCBI. A complete list of the mammalian taxa used and the accession number of the sequences is provided as Supplementary Material-4. Orthologous sequences were aligned by codons using ClustalW (the alignments can be retrieved from <http://mecom.hval.es>).

Codon sorting

Using the sequence from *Bos taurus* as reference and the crystal structure of bovine COX (Protein Data Bank, PDB, 2OCC), each codon position from the above described alignments, was sorted into different subsets according to the algorithm sketched in Figure 1. Briefly, the data set corresponding to all the codons from the alignment of a given COX subunit (for instance, chain A, corresponding to COX 1, which is a mtDNA-encoded subunit), was initially divided into two subsets: Contact and Noncontact, depending on whether the encoded amino acid from chain A, is or not closer than 4 Å to a residue from any polypeptide other than chain A, respectively. Interacting positions were defined as being < 4 Å apart because this is the upper limit for weak interactions (Martin et al. 1997). Afterwards, the Contact set was, in turn, split into two subsets: Intergenomic Contact (Mt-nu Contact, in the example) and Intragenomic Contact (Mt-mt Contact, in the example). The criterion to assign a given codon into the former subset was that the interacting residues should be encoded by different genomes, otherwise the codon is allocated into the latter subset. On the other hand, the Noncontact set was split up into two subsets: Exposed Noncontact and Buried Noncontact, on the basis of solvent accessible surface areas of the considered residue (Aledo et al. 2012). Computation of physical distances between residues and the sorting procedure was assisted by an *ad hoc* computer program that

we have called MECOM (Molecular Evolution of protein COMplexes), which, together with the corresponding manual, can be downloaded from <http://mecom.hval.es>.

Nonsynonymous sequence divergences and statistics

Nonsynonymous sequence divergences were calculated from pairwise comparisons using the alignment data subsets described in the precedent section and employing the method of maximum likelihood implemented in PAML 4.6 (Yang and Nielsen 2000; Yang 2007). The sum of these nonsynonymous sequence divergences for all the pairwise comparisons was computed and denoted as ΣdN_i , where the subscript, i , refers to the subset used for the calculations (Fig. 1).

Since we aimed to test for statistical evidence of the existence of differential evolutionary patterns among residues belonging to different subsets, we used the so-called interaction ratio (Schmidt et al. 2001), R , defined as the ratio between the ΣdN_i for the sets being tested. For instance, if we want to compare residues involved in intersubunit contacts (subset 1) with the rest of residues, which are not implicated in such interactions (subset 2), we compute $R_{12} = \Sigma dN_i / \Sigma dN_i$. In this way, a R_{12} value indistinguishable from 1 should be interpreted as that both set of residues show similar nonsynonymous substitution rates. $R_{12} < 1$ would result from a reduced evolutionary rate among the residues belonging to subset 1, relative to those residues from subset 2. On the contrary, $R_{12} > 1$ points to a higher rate of nonsynonymous substitutions among residues from subset 1, with respect to those belonging to subset 2. To determine when R_{ij} differs from 1 because of a differential pattern of evolution and when it may be due to statistical uncertainty, a Z-test was performed to assess the significance of these deviations.

For mtDNA-encoded subunits (chains A, B and C corresponding to COX1, COX2 and COX3, respectively), which are the larger subunits from complex IV (514, 227 and 261

1
2
3 436 residues, respectively), it was possible to generate a reliable random distribution of ΣdN_i . To
4
5 437 this end, the codons from multiple sequence alignments of each chain were randomly sorted
6
7 438 into two subsets of the same sizes than the original subset i . Afterwards, ΣdN_i was computed as
8
9 439 explained above. The random resampling was performed 10^4 times to build up an empirical
10
11 440 distribution.

12
13
14
15 441 In order to calculate the number of nonsynonymous substitutions per nonsynonymous
16
17 442 site on a lineage-by-lineage basis, we used a maximum likelihood method (F3x4 model)
18
19 443 implemented in codeml from the PAML package (Yang 2007). To this end, the phylogenetic
20
21 444 tree of seven mammalian species (*Bos taurus*, *Sus scrofa*, *Equus caballus*, *Ailuropoda*
22
23 445 *melanoleuca*, *Mus musculus*, *Rattus norvegicus* and *Cavia porcellus*) for which the gene
24
25 446 sequences of the 13 COX subunits were available, was reconstructed using *Arbacia lixula* as an
26
27 447 outgroup. This tree and the alignments corresponding to the different residue categories were
28
29 448 used to apportion the nonsynonymous substitutions among lineages.

30
31
32
33 449 **Relative evolutionary rate**

34
35
36 450 The codon alignment for each subunit was subjected to MEGA5 (Tamura et al. 2011)
37
38 451 to calculate the relative evolutionary rate, RER. This software implements a maximum
39
40 452 likelihood method and makes use of the Jones-Taylor-Thornton substitution model (Jones et
41
42 453 al. 1992) to estimate the relative evolutionary rate at each single site. A discrete Gamma (+G)
43
44 454 distribution was used to model evolutionary rate differences among sites. Five discrete
45
46 455 Gamma categories were considered. The computed rates were scaled such that the average
47
48 456 evolutionary rate across all sites is 1. This means that sites showing a rate lower than 1 are
49
50 457 evolving slower than average, and those with a rate higher than 1 are evolving faster than
51
52 458 average.

53
54
55
56 459 **Thermodynamic stability changes**

The thermodynamic stability changes, $\Delta\Delta G$, of mutations were computed using the protein design tool FoldX version 3.0 (Guerois et al. 2002; Schymkowitz et al. 2005). FoldX uses a full atomic description of the structure of the protein, to provide a quantitative estimation of the importance of the interactions contributing to the stability of this protein. The 3D structure of COX was subjected to an optimization procedure using the repair function of FoldX. Afterwards, an alanine scan was carried out, the resulting $\Delta\Delta G$ were recorded and used to calculate the means for Exposed Noncontact, Mt-mt Contact and Mt-nu Contact residues.

Acknowledgments

This work was supported by grant CGL2010-18124 from the Ministerio de Ciencia e Innovación, Spain. We thank Alicia Esteban del Valle and Miguel Ángel Medina for their helpful comments on the manuscript.

References

- Acín-Pérez R, Fernández-Silva P, Peleato ML, Pérez-Martos A, Enríquez JA. 2008. Respiratory Active Mitochondrial Supercomplexes. *Mol Cell* 32:529–539.
- Alba MM, Catresana J. 2005. Inverse Relationship Between Evolutionary Rate and Age of Mammalian Genes. *Mol Biol Evol* 22:598–606.
- Aledo JC. 2004. Glutamine breakdown in rapidly dividing cells: waste or investment? *Bioessays* 26:778–785.
- Aledo JC, Li Y, de Magalhães JP, Ruíz-Camacho M, Pérez-Claros JA. 2010. Mitochondrially encoded methionine is inversely related to longevity in mammals. *Aging Cell* 10:198–207.
- Aledo JC, Valverde H, Ruíz-Camacho M. 2012. Thermodynamic Stability Explains the Differential Evolutionary Dynamics of Cytochrome b and COX I in Mammals. *J Mol Evol* 74:69–80.
- Althoff T, Mills DJ, Popot J-L, Kühlbrandt W. 2011. Arrangement of electron transport chain

1
2
3 487 components in bovine mitochondrial supercomplex I1III2IV1. *EMBO J* 30:4652–4664.
4
5 488 Arnold S, Goglia F, Kadenbach B. 1998. 3,5-Diiodothyronine binds to subunit Va of
6 489 cytochrome-c oxidase and abolishes the allosteric inhibition of respiration by ATP. *Eur J*
7 490 *Biochem* 252:325–330.
8
9 491 Azevedo L, Carneiro J, van Asch B, Moleirinho A, Pereira F, Amorim A. 2009. Epistatic
10 492 interactions modulate the evolution of mammalian mitochondrial respiratory complex
11 493 components. *BMC Genomics* 10:266.
12
13 494 Barrientos A, Müller S, Dey R, Wienberg J, Morales CT (2000) Cytochrome c oxidase
14 495 assembly in primates is sensitive to small evolutionary variations in amino acid sequence.
15 496 *Mol. Biol. Evol.* 17, 1508–1519.
16
17 497 Castellana S, Vicario S, Saccone C. 2011. Evolutionary Patterns of the Mitochondrial
18 498 Genome in Metazoa: Exploring the Role of Mutation and Selection in Mitochondrial
19 499 Protein-Coding Genes. *Genome Biol Evol* 3:1067–1079.
20
21 500 Castresana J, Lübben M, Sarste M, Higgins DG. 1994. Evolution of cytochrome oxidase, an
22 501 enzyme older than atmospheric oxygen. *EMBO J* 13:2516–2525.
23
24 502 Das J, Miller ST, Stern DL. 2004. Comparison of Diverse Protein Sequences of the Nuclear-
25 503 Encoded Subunits of Cytochrome C Oxidase Suggests Conservation of Structure
26 504 Underlies Evolving Functional Sites. *Mol Biol Evol* 21:1572–1582.
27
28 505 Dudkina NV, Kudryashev M, Stahlberg H, Boekema EJ. 2011. Interaction of complexes I, III,
29 506 and IV within the bovine respirasome by single particle cryoelectron tomography. *Proc.*
30 507 *Natl Acad Sci USA* 108:15196–15200.
31
32 508 Edmands S & Burton RS. 1999. Cytochrom c oxidase activity in interpopulation hybrids of a
33 509 marine copepod: a test for nuclear-nuclear or nuclear-cytoplasmic coadaptation. *Evolution*
34 510 53:1972–1978.
35
36 511 Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. 2003. ConSurf:
37 512 Identification of functional regions in proteins by surface-Mapping of phylogenetic
38 513 information. *Bioinformatics* 19:163–164.
39
40 514 Guerois R, Nielsen JE & Serrano L. 2002. Predicting Changes in the Stability of Proteins and
41 515 Protein Complexes: A Study of More Than 1000 Mutations. *J Mol Biol* 320:369–387.
42
43 516 Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices
44 517 from protein sequences. *Computer Applications in the Biosciences* 8:275–282.
45
46 518 Kenyon L, Moraes CT. 1997. Expanding the functional human mitochondrial DNA database
47 519 by the establishment of primate xenomitochondrial cybrids. *Proc Natl Acad Sci USA*
48 520 94:9131–9135.
49
50 521 Little AG, Kocha KM, Loughheed SC, Moyes CD. 2010. Evolution of the nuclear-encoded
51 522 cytochrome oxidase subunits in vertebrates. *Physiological Genomics* 42:76–84.
52
53 523 de Magalhães JP. 2005. Human Disease-Associated Mitochondrial Mutations Fixed in
54 524 Nonhuman Primates. *J Mol Evol* 61:491–497.
55
56
57
58
59
60

- 525 Martin PD, Michael GM, Box J, Esmon CT, Edwards BF. 1997. New insights into the
526 regulation of the blood clotting cascade derived from the X-ray crystal structure of bovine
527 meizothrombine des F1 in complex with PPACK. *Structure* 5:1681–1693.
- 528 McKenzie M. 2003. Functional Respiratory Chain Analyses in Murid Xenomitochondrial
529 Cybrids Expose Coevolutionary Constraints of Cytochrome b and Nuclear Subunits of
530 Complex III. *Mol Biol Evol* 20:1117–1124.
- 531 Michel H, Behr J, Harrenga A, Kannt A. 1998. Cytochrome c Oxidase: Structure and
532 Spectroscopy. *Annu Rev Biophys Biomol Struct* 27:329–356.
- 533 Osada N, Akashi H. 2011. Mitochondrial-Nuclear Interactions and Accelerated Compensatory
534 Evolution: Evidence from the Primate Cytochrome c Oxidase Complex. *Mol Biol Evol*
535 29:337–346.
- 536 Schagger H, Pfeiffer K. 2000. Supercomplexes in the respiratory chains of yeast and
537 mammalian mitochondria. *EMBO J* 19:1777–1783.
- 538 Schagger H, Pfeiffer K. 2001. The ratio of oxidative phosphorylation complexes I-V in
539 bovine heart mitochondria and the composition of respiratory chain supercomplexes. *J*
540 *Biol Chem* 276:37861–37867.
- 541 Schmidt TR, Wildman DE, Uddin M, Opazo JC, Goodman M & Grossman LI. 2005. Rapid
542 electrostatic evolution at the binding site for cytochrome c on cytochrome c oxidase in
543 anthropoid primates. *Proc Natl Acad Sci USA* 102:6379–6384.
- 544 Schmidt TR, Wu W, Goodman M, Grossman LI. 2001. Evolution of nuclear- and
545 mitochondrial-encoded subunit interaction in cytochrome c oxidase. *Molecular Biology*
546 *and Evolution* 18:563–569.
- 547 Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F & Serrano L. 2005. The FoldX web
548 server: an online force field. *Nucleic Acids Research* 33:W382–W388.
- 549 Soto IC, Fontanesi F, Liu J & Barrientos A. 2012. Biogenesis and assembly of eukaryotic
550 cytochrome c oxidase catalytic core. *Biochimica et Biophysica Acta (BBA) -*
551 *Bioenergetics* 1817:883–897.
- 552 Suen DF, Norris KL, Youle RJ. 2008. Mitochondrial dynamics and apoptosis. *Genes &*
553 *Development* 22:1577–1590.
- 554 Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular
555 Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and
556 Maximum Parsimony Methods. *Mol Biol Evol* 28:2731–2739.
- 557 Uddin M, Opazo JC, Wildman DE, Sherwood CC, Hof PR, Goodman M, Grossman LI. 2008.
558 Molecular evolution of the cytochrome c oxidase subunit 5A gene in primates. *BMC Evol*
559 *Biol* 8:8.
- 560 Vishnoi A, Kryazhimskiy S, Bazykin GA, Hannenhalli S, Plotkin JB. 2010. Young proteins
561 experience more variable selection pressures than old proteins. *Genome Res* 20:1574–
562 1581.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

563 Willett CS. 2003. Evolution of Interacting Proteins in the Mitochondrial Electron Transport
564 System in a Marine Copepod. *Mol Biol Evol* 21:443–453.

565 Winge DR. 2012. Sealing the Mitochondrial Respirasome. *Mol Cell Biol* 32:2647–2652.

566 Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol* 24:
567 1586–1591.

568 Yang Z, Nielsen R. 2000. Estimating synonymous and nonsynonymous substitution rates
569 under realistic evolutionary models. *Mol Biol Evol* 17:32–43.

570 Yoshikawa S. 1998. Redox-Coupled Crystal Structural Changes in Bovine Heart Cytochrome
571 c Oxidase. *Science* 280:1723–1729.

572 Zhang J. 2005. Evaluation of an Improved Branch-Site Likelihood Method for Detecting
573 Positive Selection at the Molecular Level. *Mol Biol Evol* 22:2472–2479.

574

575

576

577

578

579

580

581

582

583

584

585

586

587

Table 1. Contact residues from mtDNA-encoded subunits evolve differentially depending on the genetic origin of the contacted residue.

	<i>ΣdN</i>			
	ABC	COX 1	COX 2	COX 3
Exposed Noncontact	5954 ± 10	2901 ± 9	10605 ± 104	8404 ± 50
(Mt-nu) Contact	7050 ± 15	7603 ± 40	7129 ± 56	6450 ± 44
(Mt-mt) Contact	2508 ± 7	939 ± 5	3034 ± 27	5820 ± 83
Buried Noncontact	1968 ± 3	1252 ± 4	5586 ± 70	1230 ± 14

	Relative Evolutionary Rate			
	ABC	COX 1	COX 2	COX 3
Exposed Noncontact	1.27 ± 2.26	1.03 ± 2.17	1.52 ± 1.87	1.48 ± 2.54
(Mt-nu) Contact	1.60 ± 2.33	2.33 ± 2.98	1.21 ± 1.67	0.90 ± 1.26
(Mt-mt) Contact	0.41 ± 0.92	0.26 ± 0.31	0.32 ± 0.45	0.88 ± 1.82
Buried Noncontact	0.41 ± 0.62	0.39 ± 0.53	0.78 ± 1.00	0.16 ± 0.22

590

591 In the upper table, the rates of nonsynonymous substitution per nonsynonymous site, dN,
 592 were calculated from pairwise comparisons using the alignment data subsets indicated. The
 593 sum of these nonsynonymous sequence divergences for all the pairwise comparisons was
 594 computed and denoted as *ΣdN*. Data are expressed as mean ± standard deviation. In the lower
 595 table, a different approach was used to assess the relative evolvability of the different residue
 596 subsets. The relative evolutionary rate of each residue was computed using MEGA 5. These
 597 rates were scaled such that the average evolutionary rate across all sites is 1. This means that
 598 sites showing a rate lower than 1 are evolving slower than average, and those with a rate
 599 higher than 1 are evolving faster than average. A short movie showing multiple views of the
 600 protein surface formed by residues engaged in Mt-nu (magenta) and Mt-mt contacts (blue), as

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

well as the surface of those residues that are not involved at all in intersubunit contacts (green), can be seen in Supplementary Material-5.

Table 2. Thermodynamic stability changes.

	$\Delta\Delta G$ (kJ/mol)			
	ABC	COX 1	COX 2	COX 3
Exposed Noncontact	1.41 ± 1.51 (†,§§)	1.64 ± 1.69 (§)	1.20 ± 1.09 (†,§§)	1.16 ± 1.38 (§)
Mt-mt Contact	1.78 ± 1.97 (*)	1.98 ± 1.65	1.92 ± 2.61	1.06 ± 1.52 (§)
Mt-nu Contact	1.92 ± 1.60 (**)	2.03 ± 1.55 (*)	1.99 ± 1.47 (**)	1.69 ± 1.73 (*,†)

The number of exposed noncontact residues from COX 1-3 were 144, 56 and 92, respectively. The number of mt-mt contact residues from COX 1-3 were 95, 54 and 37, respectively. The number of mt-nu contact residues from chain COX 1-3 were 120, 74 and 82, respectively. Data are expressed as mean \pm standard error.

* Significantly different from Exposed Noncontact, (*) $p < 0.05$, (**) $p < 0.0005$.

† Significantly different from Mt-mt Contact, (†) $p < 0.05$, (††) $p < 0.0005$.

§ Significantly different from Mt-nu Contact, (§) $p < 0.05$, (§§) $p < 0.0005$.

Figure legends

Figure 1. Flowchart for the main methodological procedure adopted. Once sequence and structural data were collected, aligned codons were sorted into different subsets according to the criteria sketched in the figure (details are given in the text). Afterwards, pairwise nonsynonymous sequence divergences were calculated for each subset using PAML. The sum for all comparisons yields ΣdN_i , where the subscript i denotes the subset. In addition, the mean relative evolutionary rate (RER) for each subset was computed using MEGA5. Finally, to assess whether the values of these variables were significantly different between subsets, suitable statistical tests, which are described through the text, were carried out.

Figure 2. Structural view of mitochondrially encoded COX residues. COX core, consisting of COX subunits 1 (chain A in green), 2 (chain B in yellow) and 3 (chain C in orange) is shown on the top of the figure. The spatial distribution of those residues in close contact with nDNA-encoded subunits ($ABC_{Mt_nu_Contact}$) is also shown. The set formed by mtDNA-encoded residues that are not in contact with nDNA-encoded subunits, $(ABC_{Mt_nu_Contact})^c$, was partitioned into three disjoint subsets: $ABC_{Mt_mt_Contact}$, which is formed by those residues contacting only with other mtDNA-encoded residues; $ABC_{Exposed_Noncontact}$, encompassing residues accessible to the solvent that are not involved in intersubunit contacts; and $ABC_{Buried_Noncontact}$, which contains all those residues that being buried inside the protein are not available for intersubunit contacts. The spatial distributions of the residues belonging to each of these subsets are shown at the bottom of the figure.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

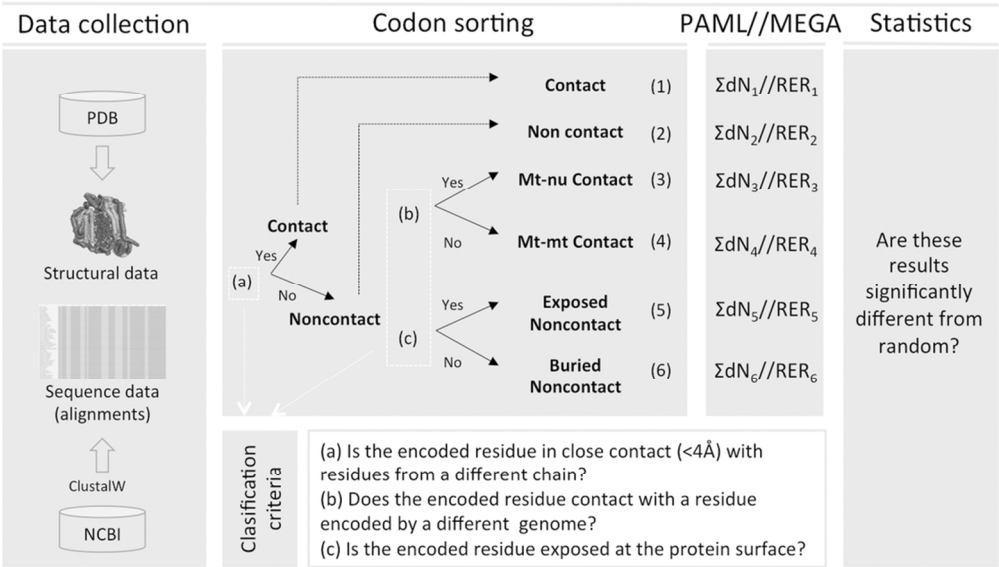
Figure 3. Patterns of nonsynonymous substitutions in COX subunits across different mammalian lineages. The phylogenetic tree of seven mammalian species (*Bos taurus*, *Sus scrofa*, *Equus caballus*, *Ailuropoda melanoleuca*, *Mus musculus*, *Rattus norvegicus* and *Cavia porcellus*) for which the gene sequences of the 13 COX subunits were available, was reconstructed using *Arbacia lixula* as an outgroup. The number of nonsynonymous substitutions per nonsynonymous site for each branch is shown for the indicated residue subset.

Figure 4. Exposed noncontact residues from COX 1 are conspicuously conserved. The Contact and Exposed Noncontact sets from each mtDNA-encoded chain were used to compute the corresponding ΣdN values. Black, light grey and dark grey bars represent COX 1, 2 and 3, respectively. From this figure, it is evident that exposed noncontact residues from COX 1 exhibit little tendency to mutate.

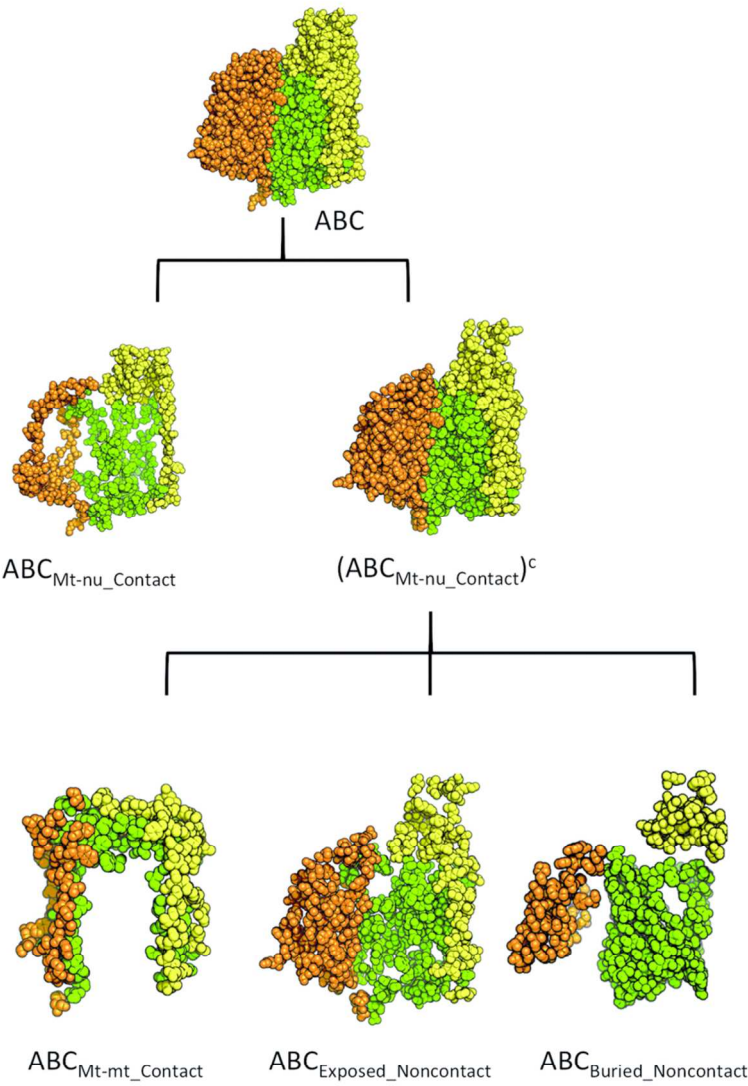
Figure 5. The behaviour of exposed noncontact residues from COX 1 diverges from that exhibited for their counterparts in COX 2 and COX 3. For each mtDNA-encoded chain the codons from the multiple sequence alignment were randomly sorted to form a subset of the same size than the original Exposed Noncontact subset of the corresponding chain. Afterwards, ΣdN was computed using this random subset. For each chain, the random resampling was performed 10^4 times to build up empirical distributions, which were used to contrast the ΣdN values computed in the real Exposed Noncontact subsets, which are indicated as filled circles on the abscissa axis.

Figure 6. Exposed noncontact residues from COX 1 are much more conserved than expected from their destabilizing effect on the holoenzyme. The exposed residues from each mtDNA-encoded proteins were split into two groups: contact (C) and noncontact (NC), according to the criteria given in the text. Afterwards, the ΣdN and mean $\Delta\Delta G$ were computed for each of these subsets. These two variables showed a significant negative correlation ($n = 6$, Pearson $r = -0.774$, p -value = 0.035) that was improved when data corresponding to COX 1 Exposed Noncontact were excluded from the analysis ($n=5$, Pearson $r = -0.890$, p -value = 0.021). The inset shows the proportion of residues that have been described as functionally relevant present among the Exposed Noncontact (grey), in comparison with the proportion of these residues that belong to the Contact group (black).

Figure 7. COX 5A, a highly conserved protein, is the only nDNA-encoded subunit with an interaction ratio R_{12} higher than the unit. (A) For each nDNA-encoded COX subunit, the ΣdN value for the Contact and Noncontact sets were computed to calculate the interaction ratio R_{12} . As it can be deduced from this figure, contact residues (magenta) tend to be more conserved than their noncontact counterpart (green), with the only exception of COX 5A. (B) To gain an indication of the relative rates of change among nuclearly encoded subunits, an alignment data set was constructed using only those species for which all the nDNA-encoded subunits were represented (*Ailuropoda melanoleuca*, *Bos taurus*, *Cavia porcellus*, *Equus caballus*, *Mus musculus*, *Rattus norvegicus* and *Sus scrofa*). Using this data set, the percentage of identity was computed for each subunit. Again COX 5A stood up, in this case as an exceptionally conserved protein.

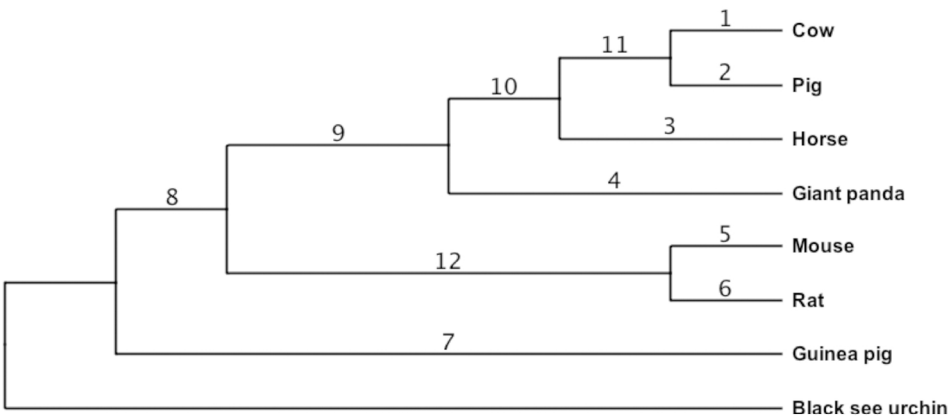


96x55mm (300 x 300 DPI)



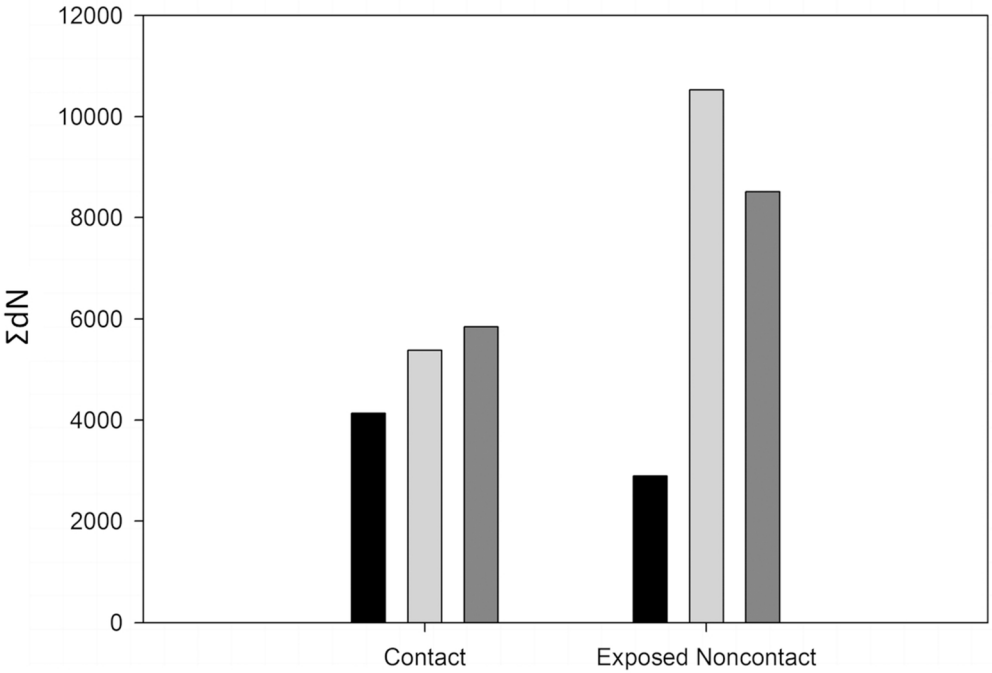
109x145mm (300 x 300 DPI)



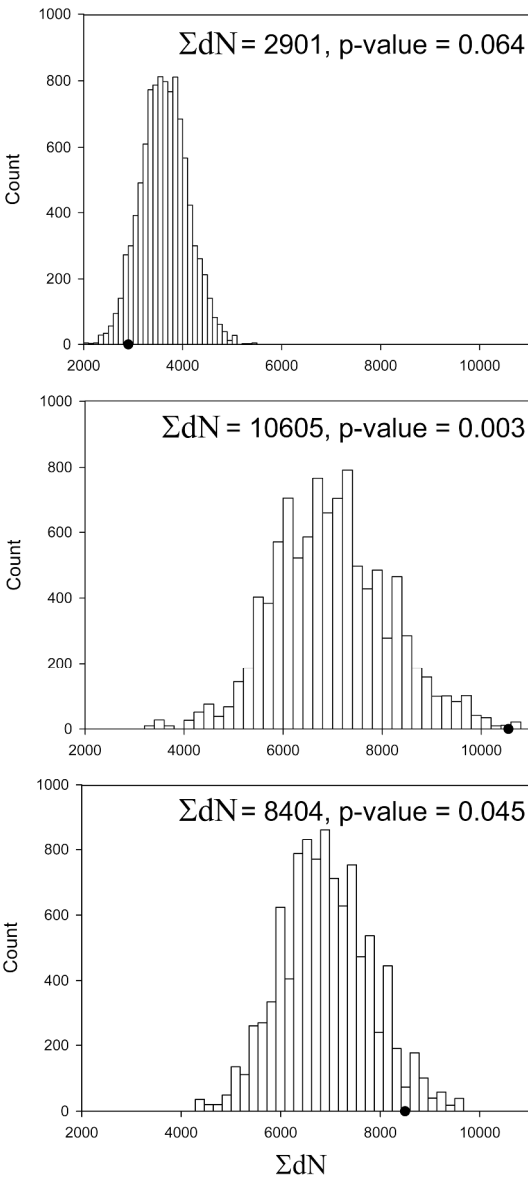


Branch	mtDNA-encoded				nDNA-encoded	
	mt-mt Contact	mt-nu Contact	Contact	Exposed NonContact	Contact	Exposed NonContact
1	0.006	0.033	0.026	0.076	0.014	0.044
2	0.006	0.046	0.035	0.037	0.011	0.030
3	0.003	0.029	0.021	0.034	0.014	0.028
4	0.006	0.046	0.033	0.090	0.021	0.028
5	0.005	0.039	0.029	0.036	0.011	0.015
6	0.003	0.020	0.015	0.034	0.012	0.389
7	0.005	0.036	0.029	0.072	0.017	0.046
8	0.005	0.038	0.027	0.068	0.017	0.035
9	0.000	0.016	0.008	0.058	0.011	0.032
10	0.003	0.020	0.011	0.000	0.002	0.005
11	0.006	0.008	0.011	0.000	0.010	0.021
12	0.006	0.062	0.044	0.119	0.034	0.064

76x71mm (600 x 600 DPI)

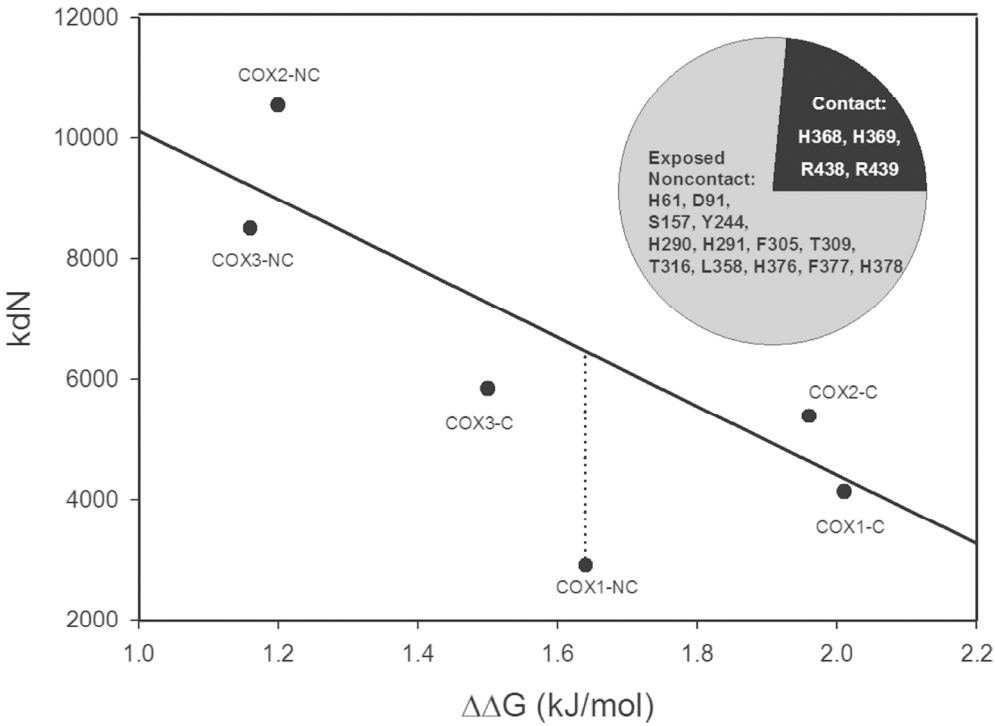


55x37mm (600 x 600 DPI)

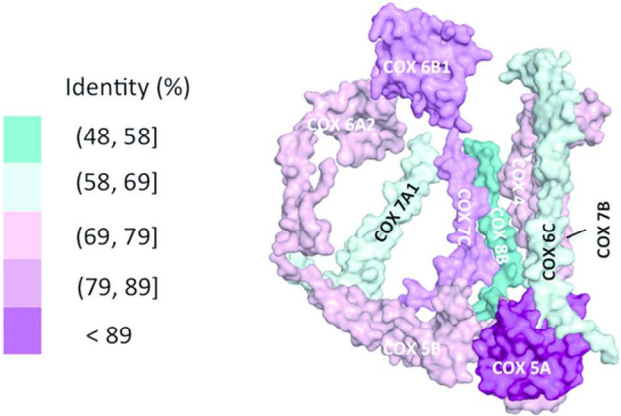
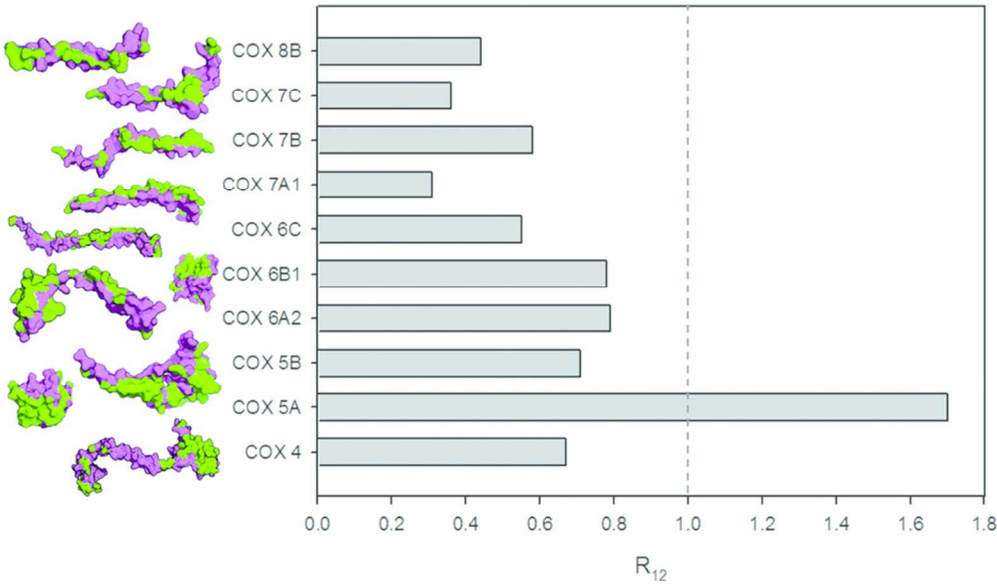


174x370mm (600 x 600 DPI)





60x44mm (600 x 600 DPI)



85x91mm (300 x 300 DPI)