New Insights into the Evolution and Gene Structure of the Mitochondrial Carrier Family Unveiled by Analyzing the Frequent and Conserved Intron Positions

Magnus Monné ,*[,1,2] Antonia Cianciulli,[1] Maria A. Panaro,[1] Rosa Calvello,[1] Anna De Grassi,[1] Luigi Palmieri ,[1,3] Vincenzo Mitolo,[1] and Ferdinando Palmieri*[,1,3]

[1]Department of Biosciences, Biotechnologies and Environment, University of Bari, Bari, Italy

[2]Department of Sciences, University of Basilicata, Potenza, Italy

[3]CNR Institute of Biomembranes, Bioenergetics and Molecular Biotechnologies (IBIOM), Bari, Italy

***Corresponding authors:** E-mails: ferdpalmieri@gmail.com; magnus.monne@unibas.it.

Associate editor: Meredith Yeager

Abstract

Mitochondrial carriers (MCs) belong to a eukaryotic protein family of transporters that in higher organisms is called the solute carrier family 25 (SLC25). All MCs have characteristic triplicated sequence repeats forming a 3-fold symmetrical structure of a six-transmembrane α-helix bundle with a centrally located substrate-binding site. Biochemical characterization has shown that MCs altogether transport a wide variety of substrates but can be divided into subfamilies, each transporting a few specific substrates. We have investigated the intron positions in the human MC genes and their orthologs of highly diversified organisms. The results demonstrate that several intron positions are present in numerous MC sequences at the same specific points, of which some are 3-fold symmetry related. Many of these frequent intron positions are also conserved in subfamilies or in groups of subfamilies transporting similar substrates. The analyses of the frequent and conserved intron positions in MCs suggest phylogenetic relationships not only between close but also distant homologs as well as a possible involvement of the intron positions in the evolution of the substrate specificity diversification of the MC family members.

Key words: **evolution, intron, intron position, mitochondrial carrier, mitochondrial transporter, solute carrier family 25, SLC25.**

Introduction

Mitochondrial carriers (MCs) constitute a family of transport proteins (the solute carrier SLC25 family), which are found in the inner membrane of mitochondria with a few exceptions in other organellar membranes (Palmieri 2004, 2013; Palmieri et al. 2011). They transport substrates of variable size, ranging from protons, inorganic ions, and small metabolites to nucleotides and large cofactors, across the membrane. Most of these substrates have been identified by the expression, purification, reconstitution into liposomes and transport assays (EPRA) approach (Palmieri and Monné 2016). The transporter domain protein sequences of MCs have hallmark features: 3 tandem 100-residue repeats, each containing 2 transmembrane segments connected by a signature motif sequence (PX[DE]XX[KR]X[RK]XQX$_{20-30}$[ED]GXXXX[YFW][RK]G; PROSITE PS50920, PFAM PF00153, and IPR00193; Saraste and Walker 1982; Palmieri 1994). The topology model for MCs exhibits six-transmembrane segments with the signature motifs toward the matrix side of the membrane and the N- and C-termini localized in the intermembrane space (Capobianco et al. 1991; Palmieri et al. 1993; Bisaccia et al. 1994; Capobianco et al. 1995). The sequence features of MCs are reflected in the three-dimensional (3D) structures of the ADP/ATP carrier (an MC family member), which consists of 6 transmembrane α-helices (H1–H6) arranged in a 3-fold pseudosymmetric bundle that confines a central pore for substrate binding and translocation (Pebay-Peyroula et al. 2003; Ruprecht et al. 2014, 2019). The central pore is alternatively closed by the "m-gate" toward the mitochondrial matrix side of the membrane (Pebay-Peyroula et al. 2003) or by the "c-gate" toward the cytoplasmic (intermembrane space) side (Ruprecht et al. 2019). The prolines of the signature motifs bend the longitudinal axes of H1, H3, and H5, and the two following charged residues in the motif form a network of interhelical salt bridges in the closed m-gate (Pebay-Peyroula et al. 2003). The last glycines of the signature motifs mark the beginning of H2, H4, and H6, each having a negatively and a positively charged residue toward the cytoplasmic side of the pore that interact in another salt bridge network to close the c-gate (Ruprecht et al. 2019). In the translocation pore, between the two gates, residues in three specific positions in H2, H4, and H6 (contact points

Article

I, II, and III, respectively) have been proposed to participate in the substrate binding (Robinson and Kunji 2006). The presence of G[IVLM], R[QHNT], or R[DE] in contact point II on H4 correlates with MCs transporting the three main classes of substrates, that is, nucleotides, carboxylated metabolites, and amino acids, respectively. Residues in the other contact points and in their vicinity are thought to provide the more precise substrate specificity (Robinson and Kunji 2006; Marobbio et al. 2008; Palmieri et al. 2011; Monné et al. 2012; Montalvo-Acosta et al. 2021; Mavridou et al. 2022). All the mentioned sequence/structural features play important roles in the conformational changes involved in the transport mechanism of MCs. In the transport cycle of MCs, the substrate enters through the open c-gate into the pore with the m-gate closed and binds the substrate-binding site, which induces closing of the c-gate and opening of the m-gate, and exits into the matrix (Klingenberg 1979; Indiveri et al. 1994; Palmieri and Pierri 2010; Palmieri et al. 2011; Kunji et al. 2016; Palmieri 2021; Ruprecht and Kunji 2021). The MC is then ready to receive a substrate from the matrix and transport it in the opposite direction through the reversal of the conformational changes indicated above. Thus most MCs are antiporters and only a few uniporters (Monné and Palmieri 2014).

The specific sequence features of MCs have been used to identify their genes in different genomes: 53 genes are present in *Homo sapiens*, 35 in *Saccharomyces cerevisiae*, and 60 in *Arabidopsis thaliana* (Palmieri and Pierri 2010; Palmieri et al. 2011). All MC genes are found on nuclear eukaryotic chromosomes with one exception found in the genome of a virus (Monné et al. 2007). Similarly to what is found in the majority of the protein-encoding genes in many eukaryotes, most MC genes have introns. Studies of protein families other than the MC family have suggested that in many distant family members the positions of introns differ substantially, indicating that these genes originated from duplications before the insertion of introns (Rogozin et al. 2012). In contrast, the intron positions (IPs) in orthologous sequences (of protein subfamilies) from various species are usually highly conserved (Rogozin et al. 2012). These observations support the hypothesis that very early in eukaryogenesis massive intron gains occurred in ancestral diversified paralogs and the subsequent evolution of their orthologs in most lineages of eukaryotes involved loss rather than gain of introns (Rogozin et al. 2012; Wang et al. 2014). The analysis of gene structure and intron distribution in protein families may therefore be used to support phylogenetic reconstruction, which otherwise is only based on protein sequence similarities (Hartung et al. 2002; Sampedro et al. 2005; Bockwoldt et al. 2019). Previously, some vertebrate MC genes have been shown to contain conserved splice sites and intronic sequences (Calvello et al. 2016, 2019; Panaro et al. 2020, 2022). In this study, we have investigated the IPs in the transporter domains of human MCs and their orthologs from various species, and analyzed the relationship between the frequent IPs and the specific sequence/structural features of MCs, their sequence similarity, and transport function. The results suggest that the frequent IPs are found at specific positions within the MC structure and that an ancestral gene had a set of IPs, which are selectively conserved in some MC subfamilies or groups of subfamilies. These conclusions have implications for the phylogenetic relationship between MC subfamilies and the evolution of diversified substrate specificity.

Results

The Intron Positions in the Human MC Sequences, Topology Model, and 3-fold Pseudosymmetrical Structure

We focused on the IPs in the transporter domains of MCs, excluding the N- or C-terminal extensions (localized outside the three tandem repeats) and therefore also the long N-terminal regulatory domains of the few calcium-binding members of the MC family. The IPs of the 53 human SLC25 family members were mapped onto their protein sequences, and these sequences were aligned against that of the bovine mitochondrial ADP/ATP carrier (supplementary fig. S1, Supplementary Material online). Because a 3D structure has been determined for the bovine mitochondrial ADP/ATP carrier (Pebay-Peyroula et al. 2003), the IPs were also located within the secondary and 3D structure of MCs. Sequence alignments of MCs are very reliable, even for distant superfamily members with low overall sequence identity, due to the highly conserved signature motifs, which fix the positions of the sequences at the end of the odd-numbered transmembrane and matrix α-helices as well as at the beginning of the even-numbered transmembrane α-helices. Insertions or deletions in the alignment were found almost exclusively in the loops connecting the secondary structure elements. The analysis of the IPs of the human SLC25 members shows that, with the exception of SLC25A2, SLC25A51, SLC25A52, and SLC25A53 that lack introns, the transporter domain sequences of human MCs have between 2 (SLC25A44) and 12 (SLC25A50) introns, of which many are located within the transmembrane α-helices. In addition, many IPs in different MCs are located in exactly the same (aligned) positions, from here on called frequent IPs, or in their vicinity, whereas other IPs are located distantly from any other. In figure 1A, the frequent IPs were displayed on the membrane topology model of MCs to facilitate inspection. The most frequent IPs among human MCs were found in the first repeat, just after the first part of the signature motif (PX[DE]XX[KR]X[RK]XQ) at the end of H1 (in 34% of the human MCs), and in the second repeat exactly in the second glycine of the signature motif ([ED]GXXXX[YFW][RK]G) at the start of H4 (in 17% of the human MCs; fig. 1A and supplementary fig. S1, Supplementary Material online). Moreover, many of the frequent IPs are found in exact or similar positions in the 3-fold repeated sequence of MCs: in the middle of the odd- and even-numbered transmembrane α-helices

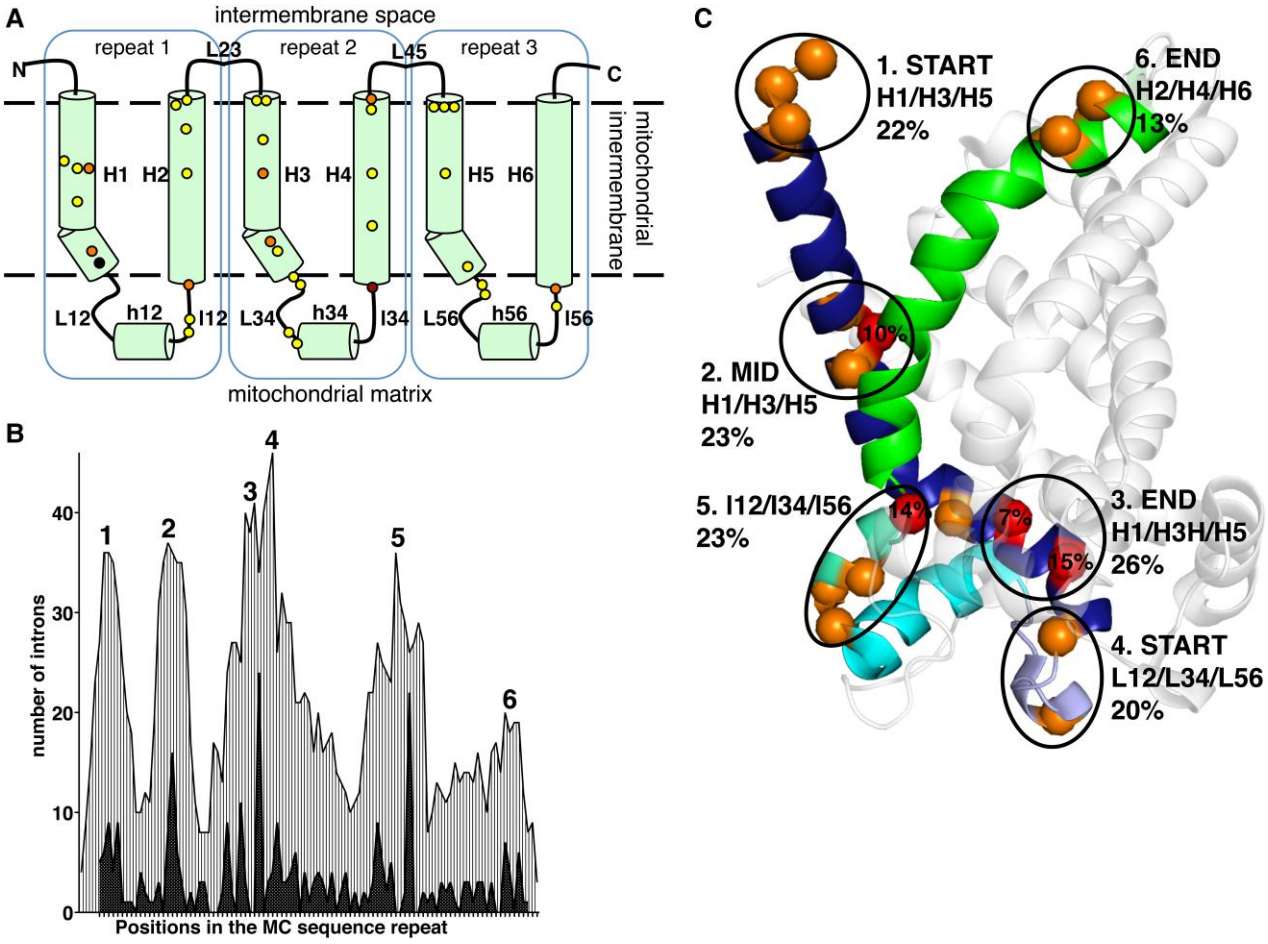

FIG. 1. Frequent IPs in human MCs analyzed in the topology model and 3-fold repeat structure. (*A*) The topology model of MCs contains six transmembrane α-helices (H1–H6), which are linked on the matrix side by loops (L12, L34, and L56), helices (h12, h34, and h56) and loops (l12, l34, and l56), and on the intermembrane space side by the loops L23 and L45. Each 100-residue sequence repeat consists of two transmembrane α-helices and their connecting segments. The most frequent IPs of the 53 human MCs are indicated by small circles colored according to the following frequency intervals: 5–10% (yellow); 10–15% (orange); 17% (dark red); and 34% (black). (*B*) The number of IPs of the three repeat sequences of all the 53 human MCs are mapped onto their 3-fold symmetry positions in the first repeat sequence of the bovine ADP/ATP carrier (consisting of H1-L12-h12-l12-H2) (dark gray graph). The numbers of IPs are summed in a 7-residue sliding window to identify aggregated IPs (maxima numbered 1–6) in the repeat sequence (light gray graph). (*C*) The frequent IPs are mapped onto the structure of the first repeat domain of the bovine ADP/ATP carrier, which is displayed in cartoon with H1 (dark blue), L12 (light blue), h12 (cyan), l12 (cyan green), and H2 (green). The rest of the ADP/ATP carrier molecule is displayed in white. The frequent IPs are indicated as C_α-spheres with frequencies 3–6% (orange) and >6% (red, with frequencies indicated), which are above the average positional frequency of about 2%. The frequent IPs are grouped into six distinct locations (encircled and corresponding to the peaks 1–6 in *B*) within the structure of the MC repeat domain: in the beginning and middle of the transmembrane α-helices H1/H3/H5 (**1** and **2**, respectively); at the end of H1/H3/H5 (**3**); in the matrix loops L12/L34/L56 (**4**); in the matrix loops l12/l35/l56 (**5**); and toward the end of the transmembrane α-helices H2/H4/H6 (**6**). The frequencies of IPs in these six locations are given in percentage.

(except for H6); toward the end of H1, H3, and H5; in the beginning and end of the matrix loops (except L12); at the end of the even-numbered α-helices (except H6); and at the beginning of the odd-numbered α-helices (except H1; fig. 1A). The latter two exceptions could be explained by the presence of introns in the N- and C-terminal extensions or in upstream and downstream noncoding sequences of their genes, that is, outside the MC transporter domain. In conclusion, these analyses show that many of the frequent IPs in human MCs are found in specific positions within the MC sequences that may be related to the signature motifs, α-helical structure, and 3-fold symmetry of these proteins.

The relationship between the IPs and the 3-fold symmetry of MCs was further analyzed by alignment of the three repeats of all human MCs on the first repeat sequence of the bovine ADP/ATP carrier. Both the exact 3-fold-symmetry-related IPs and close positions within a 7-residue proximity, which in α-helices corresponds to about one helix turn up and one helix turn down from each position, were counted (fig. 1*B*) and mapped onto the first repeat in the bovine ADP/ATP carrier structure (fig. 1C). The most frequent and exact IPs in the repeat sequences of MCs (fig. 1*A* and *B*) are found in the middle of H1/H3/H5 (in 10% of all the human MC repeats), just after the first part of the signature motif toward the end of H1/

H3/H5 (15%) and in the second glycine of the signature motif second part at the start of H2/H4/H6 (14%; fig. 1C). Closely located IPs are aggregated in six locations in the repeat sequence (fig. 1B), which in the structure correspond to: (1) the two initial helix turns of H1/H3/H5 (22% of all human MC repeats); (2) the middle of H1/H3/H5 (23%); (3) the end of H1/H3/H5 (26%); (4) the start of L12/L34/L56 (20%); (5) l1/l3/l5 (including the second glycine of the signature motif second part; 23%); and (6) toward the end of H2/H4/H6 (13%; fig. 1C). In conclusion, this analysis strengthens the hypothesis that the frequent exact and closely aggregated IPs in human MCs are found in specific locations in the 3D structure, which are related to its 3-fold pseudosymmetry.

Given that many IPs were found toward the extremities of H1–H6 (locations 1, 3, 5, and 6; fig. 1C), the IPs separating the transmembrane α-helices in MCs were re-evaluated by regarding the integrity of the core of these α-helices (the about 20 residues between the charged residues of the m- and c-gates, excluding the beginning and end of the helix). Between 74% and 78% of the human MCs have at least one intron between the core of the odd- and even-numbered transmembrane α-helices within each repeat, that is, in the matrix helices, loops L and l of each repeat (see top of fig. 2). In addition, 49% and 54% of these MCs have introns dividing the MC repeats in the cytoplasmic loops L23 and L45, respectively, that is, in and close to locations 1 and 6 (fig. 1C). In contrast to the commonly found IPs in these connecting segments, those in the core of the transmembrane α-helices H1–H6 are less common and found in 15–40% of human MCs. Furthermore, the average frequency of introns per residue in the core transmembrane α-helices is lower (ranging 0.007–0.019 [fig. 2] and on average 0.015 ± 0.003 [SE]) than that of the longer connecting segments (ranging 0.020–0.026 [fig. 2] and on average 0.024 ± 0.003) with a significant difference (Student's *t*-test, $P < 0.05$). These results show that introns in human MCs are less often found in core transmembrane α-helices than in the connecting segments.

Comparing the Intron Positions of the Human MCs

All intron-containing human MCs have at least one IP overlapping with an IP in another carrier and they may be classified into groups based on their superimposed IPs (fig. 2). Not surprisingly, many close homologs have all introns in identical positions and are therefore presented in figure 2 as one MC isoform representative, such as SLC25A4-6 (89–93% sequence identity [SI]), SLC25A7-9 (58–73% SI), SLC25A12/13 (87% SI), SLC25A14/30 (80% SI), SLC25A18/22 (66% SI), SLC25A23-25/41 (55–76% SI), SLC25A28/37 (70% SI), SLC25A33/36 (59% SI), SLC25A34/35 (51%), and SLC25A45/47 (41%). In addition, there are close homologs with only one difference in their IPs, for example, SLC25A49 and SLC25A50 (44% SI) as well as SLC25A39 and SLC25A40 (51% SI), whereas SLC25A48 has half of its six introns in positions identical to those of SLC25A45/47 (48% and 40% SI, respectively). Incidentally, the close homologs SLC25A51-53 (SLC25A51 and SLC25A52 have 97% SI and 34% SI with SLC25A53) lack introns and the intron-free SLC25A2 (not included in fig. 2) is clearly a pseudogene of SLC25A15 (88% SI) due to a quite recent evolutionary event (Fiermonte et al. 2003). In contrast, some other close MC homologs have only one IP in common, for example, SLC25A4-6 compared with SLC25A31 (71–73% SI) and SLC25A12/13 compared with SLC25A18/22 (38–40% SI). The most frequent IPs among human MCs were found in eleven positions (named A–K, of which A, E, and J are in nonexact but closely spaced positions) and may be used to further group MCs containing them (fig. 2). Especially IPs B, C, and D, which are mutually exclusive, appear to divide many of the MCs into three groups. Moreover, many carriers seem to have a combination of frequent IPs in a pattern common with others, for example, D is sometimes combined with H, and C with F, H, or I, such as in SLC25A32 (C, F, and H) and SLC25A7-9 (C, H, and K). In conclusion, many of the frequent IPs in the human carriers seem to be present and combined in specific subsets of MCs.

The frequent exact IPs (from fig. 2) were mapped onto a phylogenetic tree of the human MCs with displayed SI (>30%), substrates, and contact point II residues (fig. 3) to compare the possible groups of MCs based on these classifications. SI and phylogeny divide human MCs into four main clusters with high bootstrap values: the nucleotide carriers (MC-NT: SLC25A4-6, SLC25A16, SLC25A19, SLC25A23-25/41, SLC25A31, SLC25A43, and SLC25A42); positively charged amino acid carriers (MC-AAP: SLC25A2/15, SLC25A20 [Indiveri et al. 1997, 1998], SLC25A29 [Porcelli et al. 2014], SLC25A45/47, and SLC25A48); negatively charged amino acid carriers (MC-AAN: SLC25A12/13 and SLC25A18/22); and carboxylate carriers (MC-CA: SLC25A7-9, SLC25A10, SLC25A11, SLC25A14/30, and SLC25A34/45; fig. 3). All the other human MCs are only associated with their isoforms (through high SI) and not with any of the main clusters (bootstrap values well below 50). Not surprisingly, biochemical characterization has shown that many close homologs and isoforms of MCs transport the same or very similar substrates, such as ADP/ATP carriers (SLC25A4-6 and SLC25A31; Dolce et al. 2005), ATP-Mg/phosphate carriers (SLC25A23-25 and SLC25A41; Fiermonte et al. 2004; Traba et al. 2009), the pyrimidine nucleotide carriers (SLC25A33/36; Di Noia et al. 2014), the aspartate/glutamate carriers (SLC25A12/13; Palmieri et al. 2001), the glutamate carriers (SLC25A18/22; Fiermonte et al. 2002), the iron-transporting mitoferrins (SLC25A28/37; Shaw et al. 2006) and the carriers for sulfur oxyanions, phosphate, and dicarboxylates (SLC25A14/30; (Gorgoglione et al. 2019; fig. 3). One exception is the uncoupling protein SLC25A7 (UCP1), which transports protons (Klingenberg and Winkler 1985; Nicholls 2006), whereas SLC25A8 (UCP2) transports aspartate, malate, phosphate, and sulfate (Vozza et al. 2014). Considering

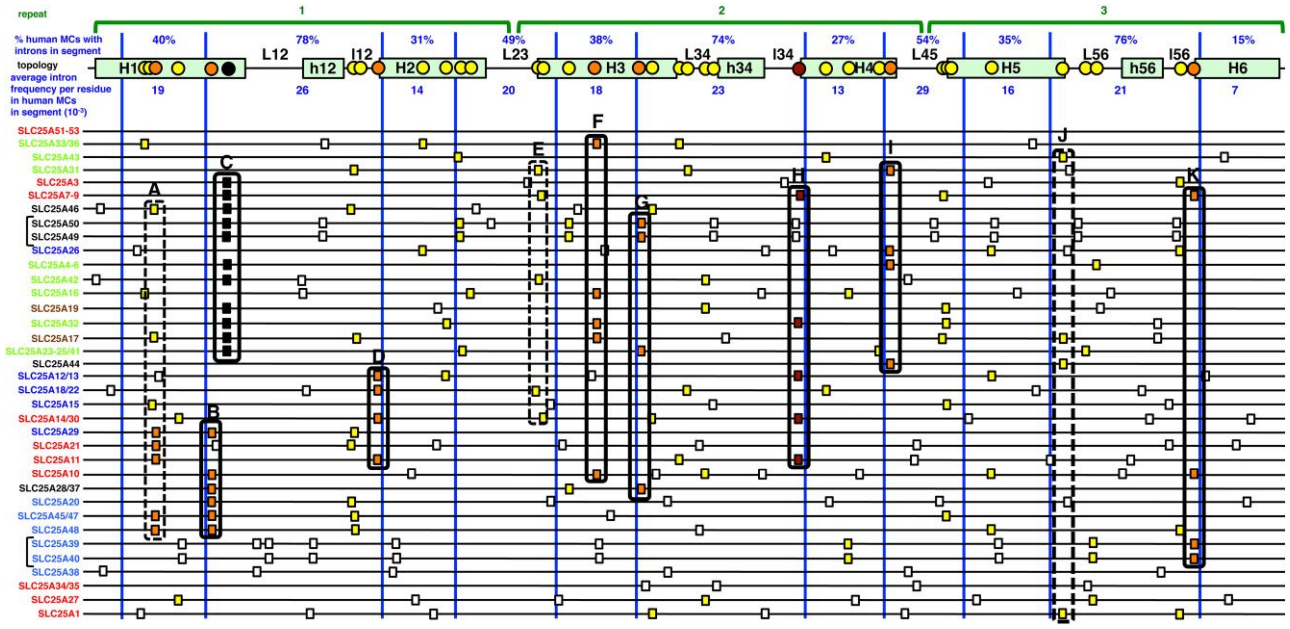

FIG. 2. IPs in the sequences of human MCs. The IPs were derived from supplementary figure S1, Supplementary Material online. The topology model of MCs (top) is displayed with the more frequent IPs as circles shown with the same colors as in figure 1A. The frequency of human MCs with at least one intron in the segments confined by the blue lines are given in percentages and the average intron frequency per residue in the segments in fractions (10^{-3}). The horizontal black lines (below) represent the human MC sequences with the IPs as small rectangles filled with the same colors as in figure 1A or with white for not-frequent IPs. The names of the MCs are colored according to their contact point II residue classification: green for G[IMTVA] (nucleotides), red for [KR]X (where X is any amino acid except D or E; carboxylates), and blue for R[DE] (amino acids). The exceptions are: SLC25A17 and SLC25A19 with LL and AI, respectively, in brown. MCs with all introns in identical positions are merged together into an isoform representative and those with almost identical IPs are grouped by brackets. The MCs are put in order to emphasize how they may be grouped according to the presence of the most frequent IPs (groups A–K), where exact IPs are encircled with solid lines and less exact positions encircled with dashed lines.

the contact point II residues, all human MCs are divided in three main classes: carriers for nucleotides, amino acids, and carboxylates, with the exception of SLC25A28/37, SLC25A44, SLC25A46, and SLC25A49-50 (fig. 3). The nucleotide carrier class, as defined by contact point II residues, correlates well with the phylogenetic MC-NT cluster and associated carriers SLC25A17, SLC25A32, and SLC25A33/36; the carboxylate carrier class correlates with the MC-CA cluster, although some carboxylate carriers are also found outside this cluster; and the amino acid carrier class comprises not only the MC-AAP and MC-AAN clusters but also MCs outside these clusters. The mapping of the frequent IPs onto the other classification systems shows that IP B is found mainly not only in the MC-AAP cluster, but also in the dicarboxylate carrier SLC25A10 (Fiermonte, Palmieri, et al. 1998) and the iron transporters SLC25A28/37. The MCs with IP C include most of the nucleotide carriers of the MC-NT cluster and associated carriers (SLC25A4-6, SLC25A23-25/41, SLC25A42 [Fiermonte et al. 2009], SLC25A19 [Rosenberg et al. 2002; Lindhurst et al. 2006], SLC25A32, and SLC25A17) as well as the phosphate carrier SLC25A3 (Fiermonte, Dolce, et al. 1998), the uncoupling proteins SLC25A7-9, the outer mitochondrial membrane protein insertases SLC25A49/50 (Guna et al. 2022) and SLC25A46, which is also found in the outer mitochondrial membrane and is involved in fusion (Abrams et al. 2015). MCs with IP D belong to the MC-AAN cluster (the asparate-glutamate carriers SLC25A12/13 and the glutamate carriers SLC25A18/22) as well as some members of the MC-CA cluster: the sulfur oxyanions/phosphate/dicarboxylate carriers SLC25A14/30 and the 2-oxoglutarate carrier SLC25A11 (Fiermonte et al. 1993). IP F is found not only in MCs for nucleotides/adenine nucleotide cofactors, such as SLC25A32 (FAD and/or folate; Titus and Moran 2000; Spaan et al. 2005), SLC25A17 (peroxisomal CoA, FAD and NAD^+; Agrimi et al. 2012), and the pyrimidine nucleotide carriers SLC25A33/36 (Di Noia et al. 2014), but also in SLC25A16 (unknown substrate) and the dicarboxylate carrier SLC25A10 (Fiermonte, Palmieri, et al. 1998). IP G is found in certain MCs, including the ATP-Mg/phosphate carriers SLC25A23-25/41, the iron transporter SLC25A28/37 and the outermembrane insertase SLC25A49/50, with no apparent similarity in substrates. MCs with IP H, that is, SLC25A7-9, SLC25A11, SLC25A12/13, SLC25A14/30, and SLC25A32, seem to combine carriers that transport malate (SLC25A8, SLC25A11, and SLC25A14/30), aspartate (SLC25A8, SLC25A12/13, and SLC25A14) and glutamate (SLC25A12/13 and SLC25A14) with FAD/folate (SLC25A32). The group with IP I combines the ADP/ATP carriers SLC25A4-6 and SLC25A31 with SLC25A26, which transports the adenosine derivatives *S*-adenosylmethionine and *S*-adenosylhomocysteine (Agrimi et al. 2004), and the hydrophobic branched chain amino

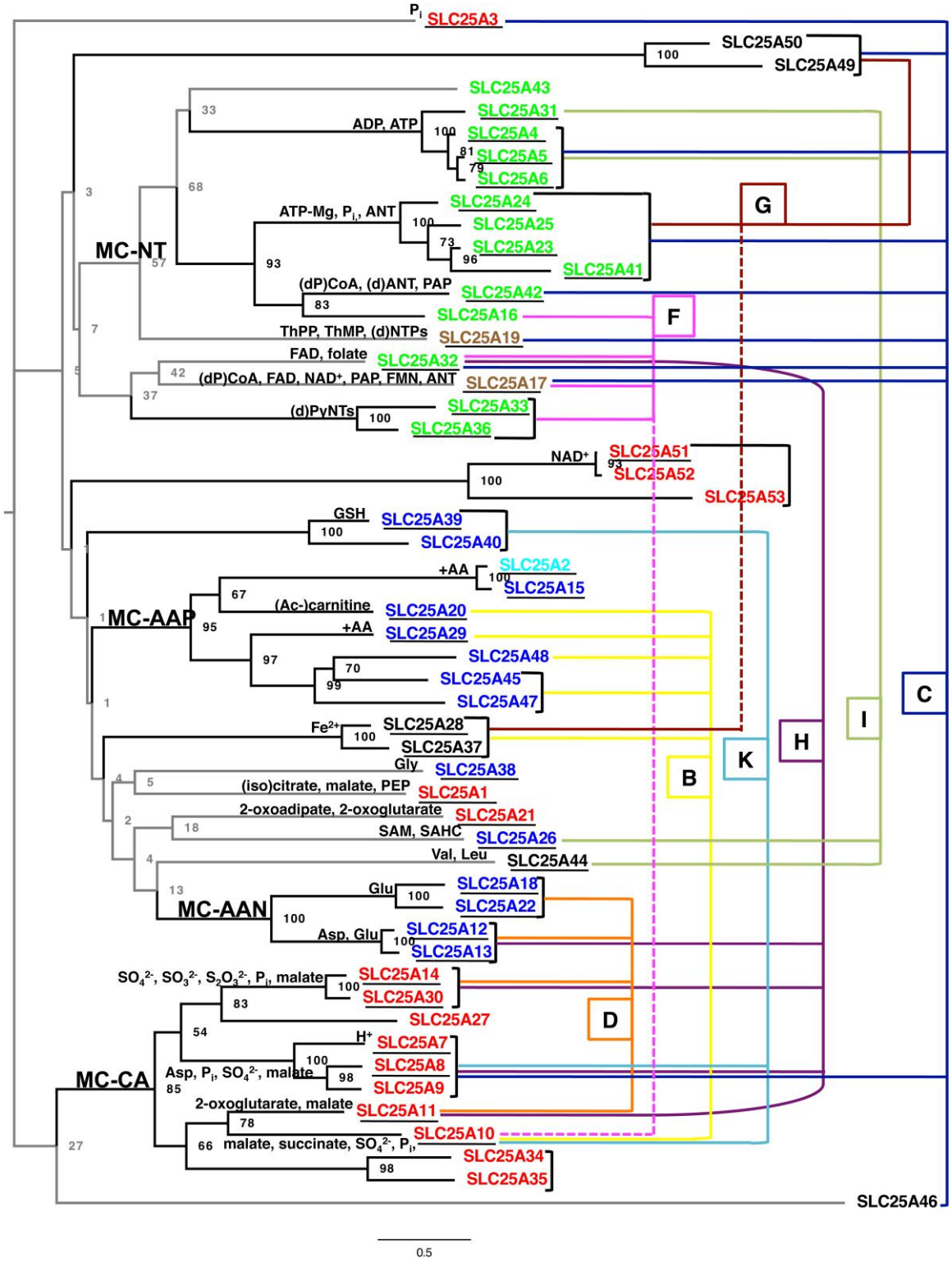

FIG. 3. Phylogenetic tree of the human MCs with their sequence identities, contact point II residue classification, main substrates, and frequent IPs indicated. The phylogenetic tree of the 53 human MCs was constructed by using PhyML v3.1 from a multiple-sequence alignment with ClustalO in Seaview4 and drawn in FigTree v1.4.2 (rooted to the left). Bootstrap values for 1,000 replicates are reported on the nodes that are connected by lines indicating sequences exhibiting <30% SI (in gray) and at least 30% SI (black). The names of the MCs are colored according to figure 2 and SLC25A2, which has QE in contact point II, in cyan. The names of MCs with known substrates are underlined and their main substrates are indicated on the nodes. The four main phylogenetic clusters of MCs for nucleotides (MC-NT), positively charged amino acids (MC-AAP), negatively charged amino acids (MC-AAN), and carboxylates (MC-CA) are indicated. The MCs that share frequent and exact IPs B, C, D, F, G, H, I, and K (from fig. 2) are linked together by colored lines on the right. +AA, positively charged amino acids; (*d*)ANT, (deoxy-)adenine nucleotides; (*d*)NTPs, (deoxy-)nucleotides; (dP)CoA, (dephospho-)coenzyme A; (*d*)PyNTs, (deoxy-)pyrimidine nucleotides; GSH, glutathione; PAP, adenosine 3′,5′-diphosphate; PEP, phosphoenolpyruvate; P_i, phosphate; SAHC, *S*-adenosylhomocysteine; SAM, *S*-adenosylmethionine; ThMP, thiamine monophosphate; ThPP, thiamine pyrophosphate.

acid carrier SLC25A44 (Yoneshiro et al. 2019). MCs with IP K include the MC-CA cluster SLC25A7-9 and SLC25A10 as well as SLC25A39/40. Among these last carriers SLC25A8 and SLC25A10 have similar substrates, for example, malate, succinate, sulfate, and phosphate, whereas SLC25A7 transports protons and SLC25A39/40 have been suggested to transport glutathione (Wang et al. 2021). In conclusion, the frequent IPs B, C, D, and F appear especially associated with MCs for positively charged amino acids, nucleotides, carboxylates/negatively charged amino acids, and nucleotide/adenine nucleotide cofactors, respectively. Therefore, the presence of the frequent IPs in MCs may be related to the evolution and substrate specificity of the members of this superfamily.

Evolutionary Conservation of Intron Positions in Human MC Orthologs

The IPs were mapped onto selected protein sequences of human MC orthologs in highly diversified representative species with confidently annotated genomes to investigate their frequency and conservation (supplementary fig. S2, Supplementary Material online). The analysis included MC sequences from five organism groups: (1) vertebrates (man, bird, amphibian, and fish; *H. sapiens*, *Gallus gallus*, *Xenopus tropicalis*, and *Danio rerio*), (2) invertebrates except insects (nematode, mollusc, tunicata, and echinodermata; *Caenorhabditis elegans*, *Aplysia californica*, *Ciona intestinalis*, and *Acanthaster planci*), (3) insects (diptera, hemiptera, lepidoptera, hymenoptera, polyneoptera, and coleoptera; *Drosophila melanogaster*, *Bemisia tabaci*, *Spodoptera frugiperda*, *Acromyrmex echinatior*, *Cryptotermes secundus*, and *Dendroctonus ponderosae*), (4) plants (eudicot, monocot, bryophyta, and green algae; *Ar. thaliana*, *Oryza sativa*, *Physcomitrium patens*, and *Chlamydomonas reinhardtii*), and (5) fungi (*Sa. cerevisiae*, *Schizosaccharomyces pombe*, *Neurospora crassa*, and *Aspergillus fumigatus*). Thus, the list includes six species of insects, as they constitute the most diversified and numerous animal group. Moreover, *Ar. thaliana* and *Sa. cerevisiae*, although all MC genes in the latter lack introns, were included because many of their MCs have been biochemically characterized (Palmieri and Monné 2016), which facilitated ortholog selection within their organism groups and the subsequent alignments. A total of 2,883 introns were found in the 654 sequences collected. From this investigation, it can be deduced that all the 211 IPs observed in the human MC isoform groups are conserved in the other vertebrates with very few exceptions. Although about 89% of the vertebrate (human) IPs are also found in at least one ortholog in the other organism groups considered, the IPs were much less conserved between the different species within each organism group compared with those of the vertebrate group. To give a simplified view of these results, the IPs of all orthologs were mapped onto the human MC sequences and displayed schematically (fig. 4*A*). About 85% of the vertebrate IPs were found in at least one noninsect invertebrate, 57% in insects, 9% in plants, and 3% in fungi. Thus, not surprisingly, the vertebrate IPs are more often conserved in other animals (the noninsect invertebrate and insect groups) than in plants and fungi. Moreover, additional IPs not found in vertebrates were identified exclusively in the other organism groups: 91 in noninsect invertebrates, 101 in insects, 140 in plants, and 51 in fungi, whereas 30 nonvertebrate IPs were found in two organism groups other than vertebrates. This means that about a third of the invertebrate IPs are uniquely found in this organism group and two-thirds are in common with vertebrates. The same figures for insects are about half unique and half shared with vertebrates, whereas about 90% of the introns in plant and fungi are unique and only 10% shared with vertebrates. Taken together, these results demonstrate that (1) the human IPs are absolutely conserved among vertebrates, (2) the majority of them are found in other animals, and (3) some of them are found in plants and fungi.

In figure 4, the frequent IPs of the human MCs were reanalyzed including the ortholog carriers from the organisms selected above on the basis of the IP frequency. In contrast to the analysis of the human MCs, the exact positions of introns A, E, and J could be defined and two new frequent IPs (G′ and I′) were added (fig. 4*A* and *B*). In general, most of the frequent IPs observed in human MCs had high frequency also among their orthologs with the exception of IPs E and I. The most frequent IP is C, which was found in 150 sequences, followed by B and J, found in 71 and 67 sequences, respectively (fig. 4*B*). In many cases, the frequent IPs are conserved (i.e., present in several organism groups) in a subset of MCs. The most conserved IPs in a single MC subfamily are C in SLC25A32 and one between G and G′ in SLC25A38, which are found in species from all five organism groups. Furthermore, many of the IPs unique for an organism group are actually found in the same position as frequent and conserved IPs in other MC subfamilies, for example, IP B (which is conserved in some MCs) is found only uniquely in plant SLC25A15 and SLC25A39/40 sequences, and IP C (which is conserved in many MCs) is found in plant SLC25A16 and SLC25A26 sequences. This phenomenon might be interpreted as the IP present in the ancestral gene (for these subfamilies) has been lost in some lineages of organisms. Therefore, this analysis suggests that most of the frequent IPs found in human MCs are partially conserved among orthologs and that the groups of MCs with specific frequent IPs may be enlarged due to their presence in orthologs.

The IPs in MC orthologs displayed in figure 4 strengthen the conclusions inferred from the analysis of the human ones regarding their relationship with the structure, phylogeny, and function of MCs. Given that the frequent MC ortholog IPs are found in the same or similar positions as the human IPs, they correlated well as far as their locations in the MC topology model and their 3-fold symmetry-related positions in the sequence repeats and structure are concerned (fig. 1). In fact, IPs A, F, and I′ are symmetry related in the middle of H1/H3/H5; B and G are in quite similar positions in H1 and H3; D, H, and

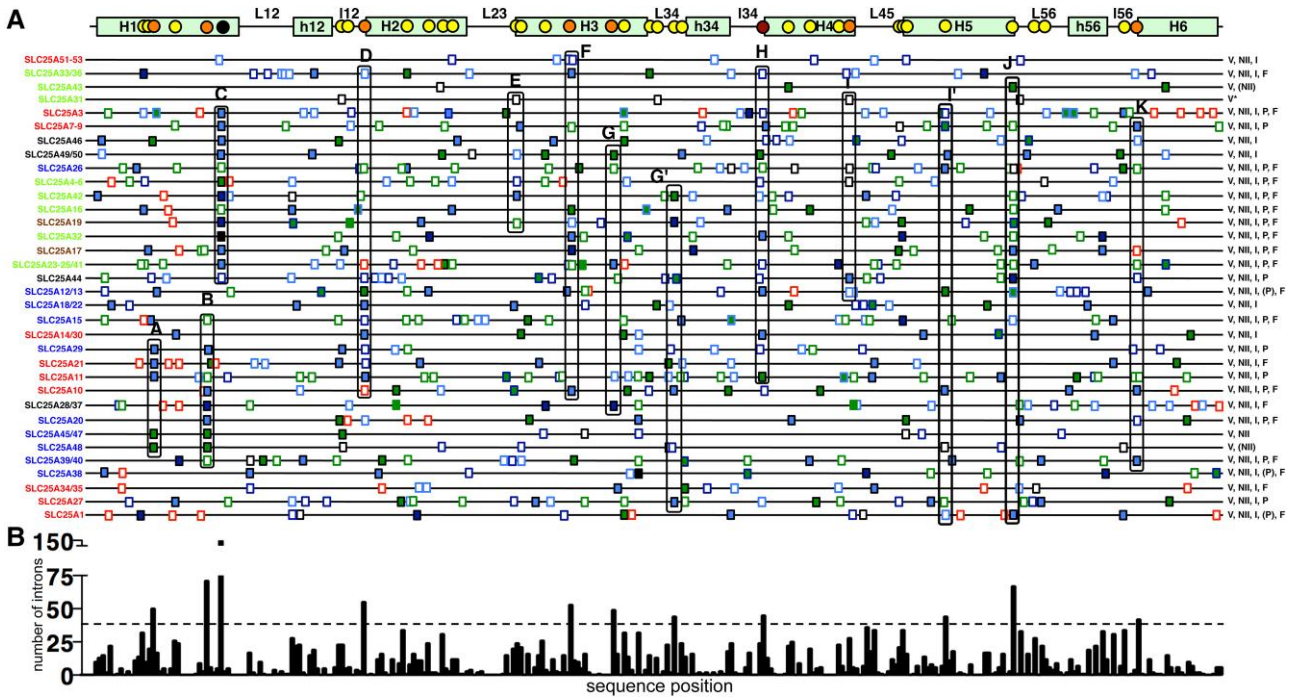

FIG. 4. IPs in human MCs and their orthologs. (*A*) The topology model of MCs (top) is displayed with the most frequent IPs in human MCs as circles shown with the same colors as in figure 1A. The horizontal black lines (below) represent the human MC sequences (their names with the color coding as in fig. 2) with the IPs of the human MCs and their orthologs represented as rectangles (as derived from supplementary fig. S2, Supplementary Material online). The IPs found exclusively in one organism group are displayed as rectangles filled white: with black frames for vertebrates (*Homo sapiens*, *Gallus gallus*, *Xenopus tropicalis*, and *Danio rerio*), with dark blue frames for noninsect invertebrates (*Caenorhabditis elegans*, *Aplysia californica*, *Ciona intestinalis*, and *Acanthaster planci*), with light blue frames for insects (*Drosophila melanogaster*, *Bemisia tabaci*, *Spodoptera frugiperda*, *Acromyrmex echinatior*, *Cryptotermes secundus*, and *Dendroctonus ponderosae*), with green frames for plants (*Arabidopsis thaliana*, *Oryza sativa*, *Physcomitrium patens*, and *Chlamydomonas reinhardtii*) and red frames for fungi (*Schizosaccharomyces pombe*, *Neurospora crassa*, and *Aspergillus fumigatus*). IPs found in more than one organism group are shown as rectangles filled with colors: in two groups (green), in three groups (light blue), in four groups (dark blue), or in all five groups (black). The most frequent and exact IPs in human MCs and their orthologs are encircled with solid lines and indicated A–K with the addition of G′ and I′. The organism group that contains orthologs of the human sequences is indicated to the right: vertebrates (V), noninsect invertebrates (NII), insects (I), plant (P), and fungi (F); and put in parenthesis if only one sequence was found in that organism group; the star means that it is only found in higher vertebrates. (*B*) The number of introns found at specific positions in the human MCs and their orthologs are plotted and aligned to the illustration in *A*. The dashed line indicates a threshold, which was chosen as four times higher (36) than the average number of introns (9) per position.

K are symmetry related at the start of H2/H4/H6 (figs. 1 and 4). In supplementary table S1, Supplementary Material online, the MCs with frequent and conserved IPs were compared with the phylogenetic analysis and biochemical function. Furthermore, a phylogenetic tree was constructed from the 654 orthologous MC sequences (supplementary fig. S3, Supplementary Material online) showing that (1) all the orthologs within each MC subfamily are clustered closely together confirming the ortholog selection; (2) all MCs within the main clusters (MC-NT, -AAP, -AAN, and -CA), as observed in the tree for the human MCs (fig. 3), are verified with only one exception represented by SLC25A34/35, which is outside the MC-CA cluster; and (3) at variance with figure 3 bootstrap values above 50 indicate that only SLC25A51-53 and SLC25A1 are significantly different than all the other MCs. The differences between the two phylogenetic trees of figure 3 and supplementary figure S3, Supplementary Material online may be explained partly by uneven diversification of ortholog sequences in specific lineages due to different evolutionary pressures and rates. Therefore, the main phylogenetic clusters as inferred from the human MCs (fig. 3) are substantially valid. Interestingly, some of the frequent MC ortholog IPs are found in MCs belonging to all or almost all four phylogenetic clusters, that is, D, H, J, and K, of which, notably, D, H, and K are symmetry related. Other IPs appear in a more limited number of groups of MCs: A is found in the MC-CA and MC-AAP clusters as well as in SLC25A21; B in the MC-CA and MC-AAP clusters, and in SLC25A28/37 and SLC25A39/40; C is a hallmark for the MC-NT cluster, its associated carriers, SLC25A7-9 and in some additional carriers (supplementary table S1, Supplementary Material online). In addition, there is a tendency for IPs co-variation (fig. 4): C is combined with the IPs E, F or G, and H or I; many MCs with D have an H IP; I is mutually exclusive with A, B, G′ and F; C and D are almost mutually exclusive. It is noteworthy that all frequent and conserved IPs are found in at least one member of the MC-CA cluster. In conclusion, the frequent IPs found in human MCs and their orthologs show that there are relationships between their locations and the MC structure, and that sets of them are not

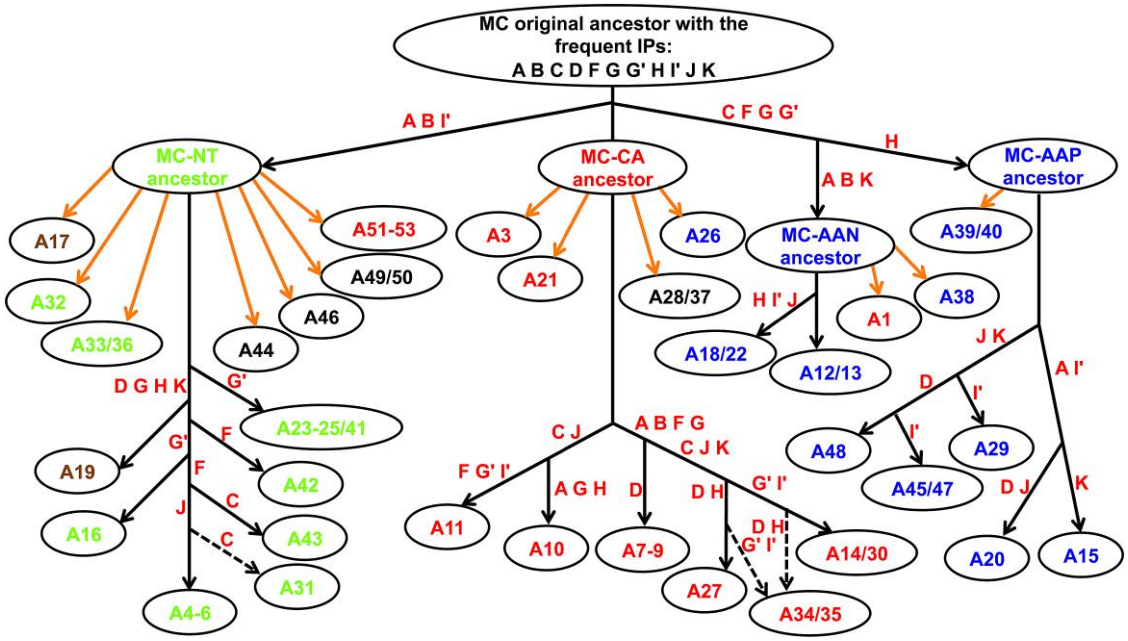

Fig. 5. Reconstructed phylogenetic tree based on the frequent and conserved IPs. The tree is based on the assumption that the total set of frequent IPs was present in the MC original ancestor followed by minimized events of repeated loss of introns. The nodes indicate the MC cluster ancestors and the ancestor for the SLCA25 subfamily orthologs (indicated with A and the specific SLCA25 number) of the nucleotide (green), carboxylate (red), and amino acid (blue) classes based on contact point II residues as in figure 2. Black arrows indicate loss of the frequent IPs indicated in red. Dashed black lines indicate alternative links. Orange arrows indicate massive loss of frequent IPs (not specified). The orange arrows originating from MC-CA ancestor could just as well have originated from the MC original ancestor.

only conserved in orthologs but also, at least partially, in different subfamilies, of which several display similarities in their substrates.

Reconstruction of an MC Superfamily Phylogenetic Tree by Using the Information of the Frequent and Conserved Intron Positions

The presence of the same frequent and conserved IPs in diverse sets of MCs from various clusters of the phylogenetic tree (figs. 3 and 4) indicates that these IPs were present early in evolution before the diversification of this family of proteins. Based on this observation, which is coherent with a scenario where the MC original ancestor contained the total or nearly total set of frequent IPs followed by subsequent partial loss of introns, a phylogenetic tree can be reconstructed through minimization of repeated intron-loss events in the branches (fig. 5). The common ancestors of each main MC cluster, therefore, possess all the IPs found among its members (supplementary table S2, Supplementary Material online). Given that the MC-CA cluster ancestor contains all frequent IPs, as they are all found in at least one of its members, whereas the MC-NT, MC-AAP, and MC-AAN ancestors contain a limited set of IPs, it is not possible to discriminate between the MC-CA ancestor and the original MC ancestor (fig. 5 and supplementary table S2, Supplementary Material online). It is also apparent that the IPs D and J cannot be used to discriminate among the ancestors of each main MC cluster, as they are found in members of all four clusters (supplementary table S2, Supplementary Material online). That said, as illustrated in the IP-reconstructed phylogenetic tree (fig. 5), the MC-NT ancestor branch deviated from the MC original/MC-CA ancestor due to the loss of IPs A, B, and I′, and the MC-AAN/MC-AAP ancestors due to the loss of IPs C, F, G, and G′ (fig. 5). Subsequently, the MC-AAN ancestor diverged because of the loss of IPs A, B, and K from the MC-AAP ancestor, which then lost H. Furthermore, as reported below, the frequent and conserved IPs in the MC subfamilies, which are found outside the main clusters, were used to link them to one of the ancestors of the four clusters, always through minimum repeated intron-loss events. In particular, the subfamily members of SLC25A3, SLC25A21, SLC25A26, and SLC25A28/37 all contain IPs that are lacking in the MC-NT, MC-AAN, and MC-AAP cluster ancestors and, therefore, most likely originated from the MC-CA ancestor or, just as likely, directly from the MC original ancestor (fig. 5 and supplementary table S2, Supplementary Material online). The ancestors of the SLC25A17, SLC25A32, SLC25A33/36, SLC25A44, SLC25A46, SLC25A49/50, and SLC25A51-53 subfamilies all derived from the MC-NT ancestor because they contain IPs not present in the MC-AAN or MC-AAP ancestors and would have lost an inferior number of frequent introns than if they would have been derived from MC-CA ancestor. Based on the same arguments SLC25A1 and SLC25A38 subfamilies derived from MC-AAN ancestor, and SLC25A39/40 subfamily from the MC-AAP ancestor. In addition, the various branches of the main cluster members developed similarly and were drawn in figure 5 also with the help of the bootstrap-supported branching of figure 3. For example, SLC25A23-25/41 orthologs contain all frequent IPs of the

MC-NT ancestor with the exception of G′, whereas all the other MC-NT members lack IPs D, G, H, and K. Although the repeated loss of IPs was minimized in the latter cluster, it was unavoidable that some IPs, such as F, G′, and C, were lost in more than one event. In the same way, the MC-CA, MC-AAN, and MC-AAP cluster members most likely diverged as illustrated in figure 5 with some repeated IP loss events. In conclusion, at variance with the previous phylogenetic trees (fig. 3 and supplementary fig. S3, Supplementary Material online), the analysis including the frequent and conserved IPs indicates that (1) SLC25A44, SLC25A46, SLC25A49/50, and SLC25A51-53 are related to the nucleotide and nucleotide-associated carriers; (2) SLC25A3, SLC25A21, SLC25A26, and SLC25A28/37 are derived from the MC-CA cluster ancestor or the original MC ancestor; (3) SLC25A1 and SLC25A38 have come from the MC-AAN cluster ancestor; and (4) SLC25A39/40 is related to the MC-AAP cluster.

Discussion

In this study, the positions of introns in the MC superfamily genes and their relationship to protein structure, function, and evolution have been analyzed. The main findings are that many of the frequent and conserved IPs in MCs are found (1) at specific positions centrally and toward the extremities of the transmembrane α-helices that are related to the 3-fold symmetry of MC sequences and structure; and (2) in specific subfamilies or groups of subfamilies with similar substrates or class of substrates. Based on these observations drawn from the IPs in MC sequences of various organisms living today, we would like to speculate on the gene architecture and the evolution of the MC superfamily members.

MCs are thought to have evolved from the triplication of a gene encoding a single 100-residue repeat; then the resulting about 300-residue ancestral MC has been multiplied into paralogs with diversified substrate specificity, which initiated the formation of MC subfamilies (Saraste and Walker 1982; Kuan and Saier 1993; Palmieri 1994; Fiermonte et al. 1999). Given that even the simple eukaryote *Sa. cerevisiae* has 35 MCs and almost all of them have orthologs among the 53 human MCs, a large part of the formation of the ancestral MC paralogs took place early in evolution, presumably during eukaryogenesis, before the fungi branch deviated from the other eukaryotes. As demonstrated in the analysis of the IPs in human MCs and their orthologs reported here, many frequent and conserved IPs are found in the same positions even in distant MC members. This observation suggests that many of the frequent IPs were already present in an ancestral MC gene and maybe already in the gene of the single 100-residue repeat. Although the positions of the intron insertions during evolution is not totally random, as they tend to be inserted in sequences that are already almost exact splice sites (proto-splice sites; Fedorov et al. 2002; Rogozin et al. 2012), it is unlikely that these introns were inserted independently and repeatedly in exactly the same corresponding positions in diversified MCs. Therefore, very probably the frequent and conserved IPs appeared before the formation of the ancestral MC paralogs and some of them were subsequently lost in certain eukaryotic lineages. It is noteworthy that remnant IPs, that could have taken part in the original triplication fusion of the repeat sequence, exist but they are less frequent and less conserved (in L23 and L45; IPs E and I). Taken together, our results point toward the possibility that the ancestral MC gene contained all the frequent and conserved IPs (A, B, C, D, F, G, G′, H, I′, J, and K as hypothesized in fig. 5), or at least IP D, H, J, and K, which are found in genes of virtually all main MC clusters. This would mean that the architecture of the original MC gene is partially conserved in the present genes of MCs.

Early on in eukaryotic evolution the intron-containing ancestral MC gene was repeatedly duplicated and diversified by mutations to form the various ancestral paralogs with different substrate specificity to provide the minimum of transport functions necessary for the early mitochondrion. Given the regularity of the frequent and conserved IPs in MCs centrally and toward the extremities of the transmembrane α-helices, exon shuffling may also have played a role in the diversification of the substrate-binding sites of the ancestral MC paralogs through exon/intron reorganization. For example, new combinations of substrate-binding sites in the primordial MC genes could have arisen by insertions/deletions or substitutions of exons encoding one of the transmembrane α-helices, an MC repeat or parts of them, which would have been facilitated by many of these exons having similar boundaries and being in the same reading frame. Alternatively or in cooperation, exon shuffling involving the IPs centrally located in the transmembrane α-helices could have led to changes of the residues in the substrate-binding site directly. Furthermore, the relocation of intron-exon boundaries of these last IPs over short distances within a range of a few nucleotides (2–3 residues) in the process called intron sliding (Rogozin et al. 2012) may also have contributed to the diversification of the substrate-binding site. During the evolution from the ancestral MC paralogs to their orthologs of the MC subfamilies in different eukaryotic species, some of the original introns were conserved but many of them were lost and a few new introns were also gained in some organism groups (as may be inferred from fig. 4). This situation is similar to what has been suggested for other protein families where the IPs often differ between distant paralogs but are conserved in orthologs (protein subfamilies; Fedorov et al. 2002; Rogozin et al. 2012; Wang et al. 2014). In conclusion, the regular arrangement of the IPs in the partially conserved gene structure of MC genes suggests that the repositioning of IPs and exon shuffling might have contributed to the diversification of the substrate specificity in MC superfamily members.

As already mentioned, some of the frequent and conserved MC IPs, especially A, B, C, and D, appear more often in certain subfamilies or groups of subfamilies with similar substrates or class of substrates. Thus, IP C is mainly found and conserved in orthologs of the MC-NT

cluster; IPs A and/or B in the MC-CA and MC-AAP clusters; and IP D in the MC-AAN and MC-CA clusters (fig. 4 and supplementary table S1, Supplementary Material online). It is noteworthy that the presence of IP C is mutually exclusive with IP A and/or B in MC genes and that all these three IPs are also found in MCs outside the four main clusters. IP C is present in orthologs of SLC25A17, SLC25A26, and SLC25A32, which transport nucleotide derivatives, and IPs A and/or B in SLC25A21 (Fiermonte et al. 2001) and SLC25A39/40 (Wang et al. 2021), which have substrates similar to the MC-CA and MC-AAP cluster members, respectively. The fact that the four frequent and conserved IPs A, B, C, and D are found mostly in the orthologs of MC subfamilies with similar substrates reflects evolutionary relationships that may help in the current attempts to identify the substrates of MCs with unknown transport function, especially of those that have very low SI with any already characterized carrier or unclear contact point II residue classification.

The reconstructed phylogenetic tree depicted in figure 5 underlines the relevance of the four main MC clusters, each characterized by human MCs related through significant bootstrap values and by the frequent and conserved IPs found in them and their orthologs. Mainly due to the grouping of MCs with the IPs A, B, C, and D, as indicated in the previous paragraph, the original ancestral MC gene is suggested to have diverged into three main branches corresponding to the MC-NT, MC-CA, and MC-AAN/MC-AAP clusters, of which the last one separated into the two well-differentiated MC-AAP and MC-AAN clusters. The IP-reconstructed phylogenetic tree also indicates that MCs outside the main clusters may be linked to one of them through minimized events of frequent IP loss. Not surprisingly, the branch with the MC-NT cluster is connected to the nucleotide(-cofactor) carriers (orthologs of SLC25A17, SLC25A32, and SLC25A33/36) and the extremely diverged NAD^+ -transporting SLC25A51-52 (Girardi et al. 2020; Kory et al. 2020; Luongo et al. 2020; Ziegler et al. 2021) (the link between the last carriers SLC25A51-53 and the MC-NT cluster is supported by the presence of the IP F in a few species, although an event of intron gain cannot be excluded in this case). The other MCs associated with the MC-NT ancestor (SLC25A44, SLC25A46, and SLC25A49/50) seem to have been derived from the MC-NT cluster (in this case, there is no apparent functional similarity, which, however, does not exclude an evolutionary link). It is noteworthy that SLC25A44, SLC25A46, SLC25A49/50, and SLC25A51-53 orthologs clearly belong to the MC superfamily in view of the presence of the frequent and conserved MC IPs, although they all lack characteristic features of this superfamily, such as many of the conserved signature motif residues, and some of them also the transport function. The subfamilies of SLC25A3, SLC25A21, SLC25A26, and SLC25A28/37 have most likely derived from the MC-CA cluster ancestor or the original MC ancestor. Among these last carriers, SLC25A21 has substrates similar to those of other MC-CA members and SLC25A3 transports phosphate, which is also transported by many of the MC-CA cluster members, whereas SLC25A26 and SLC25A28/37 have substrates different from those of the MC-CA cluster, suggesting that they may have been derived more likely from the original MC ancestor. The associations of SLC25A1 and SLC25A38 to the MC-AAN cluster, due to minimized events of repeated IP loss, are weak because they contain one and no discriminating frequent IP, respectively. However, SLC25A38 transports glycine (Lunetti et al. 2016) and, therefore, its association to the MC-AAN cluster is feasible. Another connection based on the presence of only one IP, that is, IP B, is the association of the glutathione-transporting SLC25A39/40 subfamily to the MC-AAP cluster, which, however, is supported by the similarity of the substrates they transport. In summary, the reconstruction of the phylogenetic tree based on minimized events of frequent and conserved IP loss seems to complement the evolutionary relationships established by sequence analysis rather well, that is, most close homologs (with high SI and bootstrap linking them) have many IPs in common. However, the analysis presented here also proposes evolutionary relationships between distant MC paralogs (with low SI and nonsignificant phylogenetic linkage) when frequent IPs are shared. This conclusion is substantiated by our observation that some IPs are more conserved than sequences, probably due to different evolutionary pressures on intron and coding sequences.

The reasons underlying the conservation of some IPs in the genes of specific MC subfamilies or groups of subfamilies are not yet clear. It might be that the positions and lengths of the introns are important for the positioning of different parts of the genes, such as promoters, transcription start sites, exons, or regulatory sequences, within the chromosomal architecture. Moreover, the introns of the conserved IPs may contain sequences important for gene expression regulation of MCs involved in specific metabolic pathways and cellular processes. Based only on the groups of MCs with conserved IPs in common, one might speculate, for example, that IP B contains a DNA element for regulating basic amino acid metabolism, IP C for mitochondrial oxidative phosphorylation/energy metabolism/proliferation, and IP D for the malate-aspartate shuttle/glutamate oxidation. Therefore, future studies are needed to analyze the intron sequences at the frequent and conserved IPs of MCs to associate them with regulatory elements, potential functions, and other genes.

Materials and Methods

NCBI Gene was used to collect the protein sequences of the human MCs (SLC25A1-A53) based on their SLC nomenclature. Ortholog protein sequences of the *H. sapiens* MCs and/or of the biochemically characterized MCs from *Ar. thaliana* and *Sa. cerevisiae* were searched in the NCBI nonredundant protein sequence database of the NCBI annotated genomes (of the species indicated in Results section) with BLASTP (https://blast.ncbi.nlm.nih.gov/ with

default parameters: BLOSUM62 matrix; existence and extension gap penalties of 11 and 1, respectively). The ortholog protein sequences were selected based on unambiguous reciprocal highest SI and at least 70% sequence cover with the query sequence. Multiple protein sequence alignments were done with Clustal Omega in Seaview4 with default parameters (Gouy et al. 2010; Sievers et al. 2011). Genome Data Viewer (in NCBI Gene) was used to visualize the exons and introns in "transcription-translation aligned" sequences of the genomic DNA, cDNA and protein of each of the human MCs and their orthologs to detect the IPs, which were mapped onto the protein sequence alignments manually. The IPs in the protein sequence alignments were carefully transferred into the schematic figures maintaining their proportional relative positions. The phylogenetic trees were constructed by using the protein alignments with PhyML v3.1 (Guindon et al. 2010) in Seaview4 with bootstrap values calculated for 1,000 or 100 replicates (and otherwise default parameters: model: LG; amino acid equilibrium frequencies: model-given; invariable sites: none; across site variation: optimized; tree searching operations: NNI; starting tree: BioNJ with optimized tree topology). The resulting phylogenetic trees were drawn in FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

Research in the authors' laboratories was supported by (1) the Italian Human ProteomeNet no. RBRN07BMCT_009 from the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR), (2) PRIN 2017 (2017PAB8EM) and PRIN 2020 (2020RRJP5L) grants (MIUR), and (3) Centre of Excellence on Comparative Genomics (CEGBA).

References

Abrams AJ, Hufnagel RB, Rebelo A, Zanna C, Patel N, Gonzalez MA, Campeanu IJ, Griffin LB, Groenewald S, Strickland AV, *et al.* 2015. Mutations in SLC25A46, encoding a UGO1-like protein, cause an optic atrophy spectrum disorder. *Nat Genet.* **47**:926–932.

Agrimi G, Di Noia MA, Marobbio CMT, Fiermonte G, Lasorsa FM, Palmieri F. 2004. Identification of the human mitochondrial S-adenosylmethionine transporter: bacterial expression, reconstitution, functional characterization and tissue distribution. *Biochem J.* **379**:183–190.

Agrimi G, Russo A, Scarcia P, Palmieri F. 2012. The human gene SLC25A17 encodes a peroxisomal transporter of coenzyme A, FAD and NAD+. *Biochem J.* **443**:241–247.

Bisaccia F, Capobianco L, Brandolin G, Palmieri F. 1994. Transmembrane topography of the mitochondrial oxoglutarate carrier assessed by peptide-specific antibodies and enzymatic cleavage. *Biochemistry* **33**:3705–3713.

Bockwoldt M, Heiland I, Fischer K. 2019. The evolution of the plastid phosphate translocator family. *Planta* **250**:245–261.

Calvello R, Cianciulli A, Mitolo V, Porro A, Panaro MA. 2019. Conservation of intronic sequences in vertebrate mitochondrial solute carrier genes (zebrafish, chicken, mouse and human). *Noncoding RNA.* **5**:4.

Calvello R, Panaro MA, Salvatore R, Mitolo V, Cianciulli A. 2016. Conservation/mutation in the splice sites of mitochondrial solute carrier genes of vertebrates. *J Mol Evol.* **83**:147–155.

Capobianco L, Bisaccia F, Michel A, Sluse FE, Palmieri F. 1995. The N- and C-termini of the tricarboxylate carrier are exposed to the cytoplasmic side of the inner mitochondrial membrane. *FEBS Lett.* **357**:297–300.

Capobianco L, Brandolin G, Palmieri F. 1991. Transmembrane topography of the mitochondrial phosphate carrier explored by peptide-specific antibodies and enzymatic digestion. *Biochemistry* **30**:4963–4969.

Di Noia MA, Todisco S, Cirigliano A, Rinaldi T, Agrimi G, Iacobazzi V, Palmieri F. 2014. The human SLC25A33 and SLC25A36 genes of solute carrier family 25 encode two mitochondrial pyrimidine nucleotide transporters. *J Biol Chem.* **289**:33137–33148.

Dolce V, Scarcia P, Iacopetta D, Palmieri F. 2005. A fourth ADP/ATP carrier isoform in man: identification, bacterial expression, functional characterization and tissue distribution. *FEBS Lett.* **579**:633–637.

Fedorov A, Merican AF, Gilbert W. 2002. Large-scale comparison of intron positions among animal, plant, and fungal genes. *Proc Natl Acad Sci U S A.* **99**:16128–16133.

Fiermonte G, De Leonardis F, Todisco S, Palmieri L, Lasorsa FM, Palmieri F. 2004. Identification of the mitochondrial ATP-Mg/Pi transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution. *J Biol Chem.* **279**:30722–30730.

Fiermonte G, Dolce V, Arrigoni R, Runswick MJ, Walker JE, Palmieri F. 1999. Organization and sequence of the gene for the human mitochondrial dicarboxylate carrier: evolution of the carrier family. *Biochem J.* **344** Pt 3:953–960.

Fiermonte G, Dolce V, David L, Santorelli FM, Dionisi-Vici C, Palmieri F, Walker JE. 2003. The mitochondrial ornithine transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution of two human isoforms. *J Biol Chem.* **278**:32778–32783.

Fiermonte G, Dolce V, Palmieri F. 1998. Expression in *Escherichia coli*, functional characterization, and tissue distribution of isoforms A and B of the phosphate carrier from bovine mitochondria. *J Biol Chem.* **273**:22782–22787.

Fiermonte G, Dolce V, Palmieri L, Ventura M, Runswick MJ, Palmieri F, Walker JE. 2001. Identification of the human mitochondrial oxodicarboxylate carrier. Bacterial expression, reconstitution, functional characterization, tissue distribution, and chromosomal location. *J Biol Chem.* **276**:8225–8230.

Fiermonte G, Palmieri L, Dolce V, Lasorsa FM, Palmieri F, Runswick MJ, Walker JE. 1998. The sequence, bacterial expression, and functional reconstitution of the rat mitochondrial dicarboxylate transporter cloned via distant homologs in yeast and *Caenorhabditis elegans*. *J Biol Chem.* **273**:24754–24759.

Fiermonte G, Palmieri L, Todisco S, Agrimi G, Palmieri F, Walker JE. 2002. Identification of the mitochondrial glutamate transporter. Bacterial expression, reconstitution, functional characterization, and tissue distribution of two human isoforms. *J Biol Chem.* **277**:19289–19294.

Fiermonte G, Paradies E, Todisco S, Marobbio CMT, Palmieri F. 2009. A novel member of solute carrier family 25 (SLC25A42) is a transporter of coenzyme A and adenosine 3′,5′-diphosphate In human mitochondria. *J Biol Chem.* **284**:18152–18159.

Fiermonte G, Walker JE, Palmieri F. 1993. Abundant bacterial expression and reconstitution of an intrinsic membrane-transport protein from bovine mitochondria. *Biochem J.* **294**:293–299.

Girardi E, Agrimi G, Goldmann U, Fiume G, Lindinger S, Sedlyarov V, Srndic I, Gürtl B, Agerer B, Kartnig F, *et al.* 2020. Epistasis-driven identification of SLC25A51 as a regulator of human mitochondrial NAD import. *Nat Commun.* **11**:6145.

Gorgoglione R, Porcelli V, Santoro A, Daddabbo L, Vozza A, Monné M, Di Noia MA, Palmieri L, Fiermonte G, Palmieri F. 2019. The human uncoupling proteins 5 and 6 (UCP5/SLC25A14 and UCP6/SLC25A30) transport sulfur oxyanions, phosphate and dicarboxylates. *Biochim Biophys Acta Bioenerg*. **1860**:724–733.

Gouy M, Guindon S, Gascuel O. 2010. Seaview version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol*. **27**:221–224.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate Maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. **59**:307–321.

Guna A, Stevens TA, Inglis AJ, Replogle JM, Esantsi TK, Muthukumar G, Shaffer KCL, Wang ML, Pogson AN, Jones JJ, *et al.* 2022. MTCH2 is a mitochondrial outer membrane protein insertase. *Science* **378**:317–322.

Hartung F, Blattner FR, Puchta H. 2002. Intron gain and loss in the evolution of the conserved eukaryotic recombination machinery. *Nucleic Acids Res*. **30**:5175–5181.

Indiveri C, Iacobazzi V, Giangregorio N, Palmieri F. 1997. The mitochondrial carnitine carrier protein: cDNA cloning, primary structure and comparison with other mitochondrial transport proteins. *Biochem J*. **321**:713–719.

Indiveri C, Iacobazzi V, Giangregorio N, Palmieri F. 1998. Bacterial overexpression, purification, and reconstitution of the carnitine/acylcarnitine carrier from rat liver mitochondria. *Biochem Biophys Res Commun*. **249**:589–594.

Indiveri C, Tonazzi A, Palmieri F. 1994. The reconstituted carnitine carrier from rat liver mitochondria: evidence for a transport mechanism different from that of the other mitochondrial translocators. *Biochim Biophys Acta*. **1189**:65–73.

Klingenberg M. 1979. The ADP, ATP shuttle of the mitochondrion. *Trends Biochem Sci*. **4**:249–252.

Klingenberg M, Winkler E. 1985. The reconstituted isolated uncoupling protein is a membrane potential driven H+ translocator. *EMBO J*. **4**:3087–3092.

Kory N, de Bos J U, van der Rijt S, Jankovic N, Güra M, Arp N, Pena IA, Prakash G, Chan SH, Kunchok T, *et al.* 2020. MCART1/SLC25A51 is required for mitochondrial NAD transport. *Sci Adv*. **6**(43): eabe5310.

Kuan J, Saier MH. 1993. The mitochondrial carrier family of transport proteins: structural, functional, and evolutionary relationships. *Crit Rev Biochem Mol Biol*. **28**:209–233.

Kunji ERS, Aleksandrova A, King MS, Majd H, Ashton VL, Cerson E, Springett R, Kibalchenko M, Tavoulari S, Crichton PG, *et al.* 2016. The transport mechanism of the mitochondrial ADP/ATP carrier. *Biochim Biophys Acta*. **1863**:2379–2393.

Lindhurst MJ, Fiermonte G, Song S, Struys E, De Leonardis F, Schwartzberg PL, Chen A, Castegna A, Verhoeven N, Mathews CK, *et al.* 2006. Knockout of Slc25a19 causes mitochondrial thiamine pyrophosphate depletion, embryonic lethality, CNS malformations, and anemia. *Proc Natl Acad Sci U S A*. **103**: 15927–15932.

Lunetti P, Damiano F, De Benedetto G, Siculella L, Pennetta A, Muto L, Paradies E, Marobbio CMT, Dolce V, Capobianco L. 2016. Characterization of human and yeast mitochondrial glycine carriers with implications for heme biosynthesis and anemia. *J Biol Chem*. **291**:19746–19759.

Luongo TS, Eller JM, Lu M-J, Niere M, Raith F, Perry C, Bornstein MR, Oliphint P, Wang L, McReynolds MR, *et al.* 2020. SLC25A51 is a mammalian mitochondrial NAD+ transporter. *Nature* **588**:174–179.

Marobbio CMT, Giannuzzi G, Paradies E, Pierri CL, Palmieri F. 2008. α-Isopropylmalate, a leucine biosynthesis intermediate in yeast, is transported by the mitochondrial oxalacetate carrier. *J Biol Chem*. **283**:28445–28553.

Mavridou V, King MS, Tavoulari S, Ruprecht JJ, Palmer SM, Kunji ERS. 2022. Substrate binding in the mitochondrial ADP/ATP carrier is a step-wise process guiding the structural changes in the transport cycle. *Nat Commun*. **13**:3585.

Monné M, Miniero DV, Daddabbo L, Robinson AJ, Kunji ERS, Palmieri F. 2012. Substrate specificity of the two mitochondrial ornithine carriers can be swapped by single mutation in substrate binding site. *J Biol Chem*. **287**:7925–7934.

Monné M, Palmieri F. 2014. Antiporters of the mitochondrial carrier family. *Curr Top Membr*. **73**:289–320.

Monné M, Robinson AJ, Boes C, Harbour ME, Fearnley IM, Kunji ERS. 2007. The mimivirus genome encodes a mitochondrial carrier that transports dATP and dTTP. *J Virol*. **81**:3181–3186.

Montalvo-Acosta JJ, Kunji ERS, Ruprecht JJ, Dehez F, Chipot C. 2021. Structure, substrate binding, and symmetry of the mitochondrial ADP/ATP carrier in its matrix-open state. *Biophys J*. **120**:5187–5195.

Nicholls DG. 2006. The physiological regulation of uncoupling proteins. *Biochim Biophys Acta*. **1757**:459–466.

Palmieri F. 1994. Mitochondrial carrier proteins. *FEBS Lett*. **346**:48–54.

Palmieri F. 2004. The mitochondrial transporter family (SLC25): physiological and pathological implications. *Pflügers Archiv*. **447**:689–709.

Palmieri F. 2013. The mitochondrial transporter family SLC25: identification, properties and physiopathology. *Mol Aspects Med*. **34**: 465–484.

Palmieri F. 2021. Mitochondrial transporters of the solute carrier 25 (SLC25) superfamily. In: Jez J, editor. *Encyclopedia of biological chemistry III. 3rd ed*. Oxford: Elsevier. p. 213–243.

Palmieri F, Bisaccia F, Capobianco L, Dolce V, Fiermonte G, Iacobazzi V, Zara V. 1993. Transmembrane topology, genes, and biogenesis of the mitochondrial phosphate and oxoglutarate carriers. *J Bioenerg Biomembr*. **25**:493–501.

Palmieri F, Monné M. 2016. Discoveries, metabolic roles and diseases of mitochondrial carriers: a review. *Biochim Biophys Acta*. **1863**: 2362–2378.

Palmieri L, Pardo B, Lasorsa FM, del Arco A, Kobayashi K, Iijima M, Runswick MJ, Walker JE, Saheki T, Satrústegui J, *et al.* 2001. Citrin and aralar1 are Ca(2+)-stimulated aspartate/glutamate transporters in mitochondria. *EMBO J*. **20**:5060–5069.

Palmieri F, Pierri CL. 2010. Mitochondrial metabolite transport. *Essays Biochem*. **47**:37–52.

Palmieri F, Pierri CL, De Grassi A, Nunes-Nesi A, Fernie AR. 2011. Evolution, structure and function of mitochondrial carriers: a review with new insights. *Plant J*. **66**:161–181.

Panaro MA, Calvello R, Miniero DV, Mitolo V, Cianciulli A. 2022. Imaging intron evolution. *Methods Protoc*. **5**:53.

Panaro MA, Calvello R, Mitolo V, Cianciulli A. 2020. 5′ And 3′ splicing signals evolution in vertebrates: analysis in a conserved gene family. *Comput Biol Chem*. **86**:107251.

Pebay-Peyroula E, Dahout-Gonzalez C, Kahn R, Trézéguet V, Lauquin GJ-M, Brandolin G. 2003. Structure of mitochondrial ADP/ATP carrier in complex with carboxyatractyloside. *Nature* **426**:39–44.

Porcelli V, Fiermonte G, Longo A, Palmieri F. 2014. The human gene SLC25A29, of solute carrier family 25, encodes a mitochondrial transporter of basic amino acids. *J Biol Chem*. **289**:13374–13384.

Robinson AJ, Kunji ERS. 2006. Mitochondrial carriers in the cytoplasmic state have a common substrate binding site. *Proc Natl Acad Sci U S A*. **103**:2617–2622.

Rogozin IB, Carmel L, Csuros M, Koonin EV. 2012. Origin and evolution of spliceosomal introns. *Biol Direct*. **7**:11.

Rosenberg MJ, Agarwala R, Bouffard G, Davis J, Fiermonte G, Hilliard MS, Koch T, Kalikin LM, Makalowska I, Morton DH, *et al.* 2002. Mutant deoxynucleotide carrier is associated with congenital microcephaly. *Nat Genet*. **32**:175–179.

Ruprecht JJ, Hellawell AM, Harding M, Crichton PG, McCoy AJ, Kunji ERS. 2014. Structures of yeast mitochondrial ADP/ATP carriers support a domain-based alternating-access transport mechanism. *Proc Natl Acad Sci U S A*. **111**:E426–E434.

Ruprecht JJ, King MS, Zogg T, Aleksandrova AA, Pardon E, Crichton PG, Steyaert J, Kunji ERS. 2019. The molecular mechanism of transport by the mitochondrial ADP/ATP carrier. *Cell* **176**: 435–447.

Ruprecht JJ, Kunji ERS. 2021. Structural mechanism of transport of mitochondrial carriers. *Annu Rev Biochem*. **90**:535–558.
Sampedro J, Lee Y, Carey RE, de Pamphilis C, Cosgrove DJ. 2005. Use of genomic history to improve phylogeny and understanding of births and deaths in a gene family. *Plant J*. **44**:409–419.
Saraste M, Walker JE. 1982. Internal sequence repeats and the path of polypeptide in mitochondrial ADP/ATP translocase. *FEBS Lett*. **144**:250–254.
Shaw GC, Cope JJ, Li L, Corson K, Hersey C, Ackermann GE, Gwynn B, Lambert AJ, Wingert RA, Traver D, *et al*. 2006. Mitoferrin is essential for erythroid iron assimilation. *Nature* **440**:96–100.
Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Söding J, *et al*. 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. **7**:539.
Spaan AN, Ijlst L, van Roermund CWT, Wijburg FA, Wanders RJA, Waterham HR. 2005. Identification of the human mitochondrial FAD transporter and its potential role in multiple acyl-CoA dehydrogenase deficiency. *Mol Genet Metab*. **86**:441–447.
Titus SA, Moran RG. 2000. Retrovirally mediated complementation of the glyB phenotype. Cloning of a human gene encoding the carrier for entry of folates into mitochondria. *J Biol Chem*. **275**:36811–36807.
Traba J, Satrústegui J, del Arco A. 2009. Characterization of SCaMC-3-like/slc25a41, a novel calcium-independent mitochondrial ATP-Mg/pi carrier. *Biochem J*. **418**:125–133.
Vozza A, Parisi G, De Leonardis F, Lasorsa FM, Castegna A, Amorese D, Marmo R, Calcagnile VM, Palmieri L, Ricquier D, *et al*. 2014. UCP2 Transports C4 metabolites out of mitochondria, regulating glucose and glutamine oxidation. *Proc Natl Acad Sci U S A*. **111**:960–965.
Wang H, Devos KM, Bennetzen JL. 2014. Recurrent loss of specific introns during angiosperm evolution. *PLoS Genet*. **10**:e1004843.
Wang Y, Yen FS, Zhu XG, Timson RC, Weber R, Xing C, Liu Y, Allwein B, Luo H, Yeh H-W, *et al*. 2021. SLC25A39 Is necessary for mitochondrial glutathione import in mammalian cells. *Nature* **599**:136–140.
Yoneshiro T, Wang Q, Tajima K, Matsushita M, Maki H, Igarashi K, Dai Z, White PJ, McGarrah RW, Ilkayeva OR, *et al*. 2019. BCAA Catabolism in brown fat controls energy homeostasis through SLC25A44. *Nature* **572**:614–619.
Ziegler M, Monné M, Nikiforov A, Agrimi G, Heiland I, Palmieri F. 2021. Welcome to the family: identification of the NAD+ transporter of animal mitochondria as member of the solute carrier family SLC25. *Biomolecules* **11**:880.