npg

EJHG*Open*

ARTICLE

Mitochondrial cardiomyopathies: how to identify candidate pathogenic mutations by mitochondrial DNA sequencing, MITOMASTER and phylogeny

Michael V Zaragoza*[1,2], Martin C Brandon[1], Marta Diegoli[3], Eloisa Arbustini[3] and Douglas C Wallace*[1,4]

Pathogenic mitochondrial DNA (mtDNA) mutations leading to mitochondrial dysfunction can cause cardiomyopathy and heart failure. Owing to a high mutation rate, mtDNA defects may occur at any nucleotide in its 16 569 bp sequence. Complete mtDNA sequencing may detect pathogenic mutations, which can be difficult to interpret because of normal ethnic/geographic-associated haplogroup variation. Our goal is to show how to identify candidate mtDNA mutations by sorting out polymorphisms using readily available online tools. The purpose of this approach is to help investigators in prioritizing mtDNA variants for functional analysis to establish pathogenicity. We analyzed complete mtDNA sequences from 29 Italian patients with mitochondrial cardiomyopathy or suspected disease. Using MITOMASTER and PhyloTree, we characterized 593 substitution variants by haplogroup and allele frequencies to identify all novel, non-haplogroup-associated variants. MITOMASTER permitted determination of each variant's location, amino acid change and evolutionary conservation. We found that 98% of variants were common or rare, haplogroup-associated variants, and thus unlikely to be primary cause in 80% of cases. Six variants were novel, non-haplogroup variants and thus possible contributors to disease etiology. Two with the greatest pathogenic potential were heteroplasmic, nonsynonymous variants: m.15132T>C in *MT-CYB* for a patient with hypertrophic dilated cardiomyopathy and m.6570G>T in *MT-CO1* for a patient with myopathy. In summary, we have used our automated information system, MITOMASTER, to make a preliminary distinction between normal mtDNA variation and pathogenic mutations in patient samples; this fast and easy approach allowed us to select the variants for traditional analysis to establish pathogenicity.

INTRODUCTION

Mitochondrial diseases result from abnormalities in the mitochondrion, the essential organelle that provides energy in the form of ATP for normal heart function.[1] Mitochondria are dynamic organelles that number from several to hundreds per cell with each mitochondrion containing multiple copies of mitochondrial DNA (mtDNA). Normal mitochondrial function relies on two types of inheritance: maternal inheritance of circular mtDNA sequences and biparental inheritance of nuclear DNA (nDNA) that has over 1000 genes for mitochondrial homeostasis. Consequently, mitochondrial dysfunction because of mtDNA defects or nDNA mutations can result in disease with cardiomyopathy and heart failure as the major features.[2]

mtDNA is the circular, 16 569 base sequence with 37 genes that encode for 13 proteins, 22 transfer RNAs (tRNAs) and 2 ribosomal RNAs.[3] Heteroplasmy describes the state of a mixture of normal and mutant copies of mtDNA. Depending on the timing of origin, location and segregation of mutant mtDNA, an individual may have different ratios of mutant and normal mtDNA distributed amongst different tissues or organs that result in disease.[4]

Owing to these genetic factors, mitochondrial diseases leading to myocardial disease have much clinical heterogeneity in age of presentation and severity of symptoms.[5] Cardiac features include concentric ventricular hypertrophy, dilated cardiomyopathy (DCM) and congestive heart failure.[6–8] Cardiac conduction features include atrioventricular and bundle branch blocks.[9] When mitochondrial disease primarily affects only the heart, hypertrophic cardiomyopathy (HCM) or dilated mitochondrial cardiomyopathies may be clinically indistinguishable from other cardiomyopathies.[6,7,10] Diagnosis may be difficult requiring cardiac tissue for diagnosis; histopathological features include changes in mitochondrial shape or number, histochemical defects and abnormities in oxidative phosphorylation (OXPHOS) enzyme activities.[5]

Diagnosis of mitochondrial cardiomyopathies may be achieved by mtDNA sequence analysis for known pathogenic mtDNA mutations. For example, testing may be carried out for the m.3243A>G mutation in the tRNA leucine gene (*MT-TL1*) associated with MELAS (mitochondrial myopathy, encephalopathy, lactic acidosis and stroke).[11] The m.3243A>G mutation may also lead to severe, concentric HCM or DCM presenting in infancy or childhood.[10,12] Adults with m.3243A>G also may present with cardiomyopathy and progressive heart failure with extracardiac features of mitochondrial disease including deafness and diabetes.[13,14]

[1]Center for Mitochondrial and Molecular Medicine and Genetics, University of California, Irvine, CA, USA; [2]Department of Biological Chemistry, University of California, Irvine, CA, USA; [3]Division of Genetics and Metabolism, Department of Pediatrics, University of California, Irvine, CA, USA; [4]The Centre for Inherited Cardiovascular Diseases, IRCCS Foundation Policlinico San Matteo, Pavia, Italy
*Correspondence: Dr MV Zaragoza, Center for Mitochondrial and Molecular Medicine and Genetics, University of California, 2011 Hewitt Hall, Irvine, CA 92697, USA. Tel: +19 49 824 8813; Fax: +19 49 824 6388; E-mail: mzaragoz@uci.edu or Dr DC Wallace, Director, Center for Mitochondrial and Molecular Medicine and Genetics, University of California, Irvine, 2026 Hewitt Hall, Irvine, CA 92697, USA. Tel: +19 49/824-3490; Fax: +19 49/824-6388; E-mail: dwallace@uci.edu

A newly arising pathogenic mtDNA mutation will be heteroplasmic. As the percentage of mutant mtDNAs increases, the severity of the clinical phenotype at each percentage mutation will depend on the severity of the biochemical defect caused by the mutation. Severe mtDNA mutations will result in reproductive failure while still heteroplasmic. Hence, surviving patients with the mutation will be heteroplasmic. By contrast, milder mtDNA mutations may not cause clinical symptoms severe enough to affect reproduction until they reach homoplasmy. These later mutations have proven difficult to distinguish from neutral or adaptive polymorphisms.

It is now standard-of-care to sequence the mtDNAs of patients with complex diseases, such as cardiomyopathy associated with mitochondrial dysfunction.[6–8,15–17] The mtDNAs of over 500 patients with HCM or DCM patients have already been sequenced and the mtDNA variants that differ from a reference sequence have been compared with small samples of 'control' cases. Those variants observed in the patient mtDNAs but not in the study controls have been considered to be potential disease causing mutations. Over 1000 variants have been reported in various cardiomyopathy cases, which encompass 200 different sequence variants.[6–8,15–17]

However, population genetic studies have showed that the 'normal' mtDNA sequence variation is very high. This is the result of ancient polymorphisms associated with distinct ethnic and/or geographic-associated maternal lineages, known as mtDNA haplogroups.[18] Distinctive haplogroup lineages are descendent from a single founding maternal ancestor mtDNAs resulting in a discrete branch of the mtDNA phylogenetic tree which shares the variants of the founding mtDNA.[19–22] Therefore, if a patient is from a rare mtDNA haplogroup, then it is unlikely the control samples used in that study will also include a mtDNA from that haplogroup. Hence, many of the haplogroup-associated normal variants will be interpreted as potentially pathogenic.

This complexity can in part be resolved by comparing the mtDNA sequence variants of a patient with those of a large database of mtDNA sequences that have been delineated by haplogroup. If the database was exhaustive, then most of haplogroup-associated variants would be represented within the database, at frequencies consistent with the haplogroup distribution within the population. Therefore, variants associated with specific populations will be both routinely linked to other variants associated with that same haplogroup and will also be present at a significant frequency within the overall population of mtDNA sequences.

At the other extreme, a recently arising pathogenic variant will not be repeatedly linked to the same array of mtDNA variants and will be very rarely in the overall population, as it is continually being removed by selection. Therefore, novel mtDNA variants or one found very rarely in the population have a greater probability of being pathogenic. If the variant also affects a functionally important mitochondrial function and is heteroplasmic, this further increases the likelihood that the mutation is contributing to the disease.[21]

On the basis of this logic, we have developed a database of several thousand mtDNA sequences that encompasses much of the global mtDNA variation. In addition, we have developed a computer program, MITOMASTER,[23] which will analyze a patient mtDNA sequence based on its deviations from a master reference sequence and then use this list of variants to deduce the patient's haplogroup. The variants are then compared with all other mtDNAs in that haplogroup and the common haplogroup variants identified. The frequency of each variant is also calculated.

Rare, non-haplogroup-associated variants are then analyzed for the gene affected, sequence conservation, and functional consequences permitting assessment of the potential pathogenicity of the variant. To test the capacity of this system to help assess the pathogenic potential of mtDNA mutations, we now apply our MITOMASTER analysis system to a set of 29 mtDNA sequences from patients with mitochondrial cardiomyopathy and other mitochondrial clinical phenotypes. The systematic approach allowed us to prioritize mtDNA sequence variants for functional analysis to establish pathogenicity.

METHODS

Study cohort

We evaluated 29 patients referred to Dr Eloisa Arbustini at the Centre for Inherited Cardiovascular Diseases (CICD) in Pavia, Italy. We obtained informed written consent of all participants in accordance with the IRCCS Foundation Policlinico San Matteo in Pavia and with the University of California (UCI), Irvine for de-identified DNA samples. This included 21 patients with highly likely mitochondrial cardiomyopathy and 8 suspected of mitochondrial disease as revealed by clinical, biochemical and histopathological findings.[8]

Criteria for suspecting matrilineal mitochondrial cardiomyopathy included clinical screening and pedigree evaluation with exclusion of families with male to offspring transmission of the disease; cardiac imaging with left ventricular (LV) dilated hypokinetic phenotypes and concentric LV hypertrophy early in the course of disease ('hypertrophic→dilated cardiomyopathies'); stable or recurrent lactic academia; clinical traits of mtDNA-related diseases (>sCPK, myopathy, encephalomyopathies, palpebral ptosis and so on) in the proband and in maternal relatives); identification of potentially pathologic mtDNA variants by whole mtDNA sequencing; when possible, preliminary assessment for homoplasmy/heteroplasmy of mtDNA variants by RFLP or Q-PCR in endomyocardial biopsies (EMB) compared with blood; mitochondrial proliferation and abnormal morphology in EMB samples; and loss of COX activity in the myocardial samples.[8]

Table 1 shows the major clinical features for all 29 cases; 19 singleton and 10 familial cases based on pedigree analysis and echocardiography of family members. Case 1 was from previously reported HCM family with an MYH7 mutation and possible mtDNA mutation m.9957T>C.[24]

Sanger sequencing of mtDNA

We extracted genomic DNA from blood except for case 7 in which we used heart tissue. It should be noted that severely deleterious mtDNA mutations that cause biochemical defects while heteroplasmic can be selectively lost from blood, even though they are present in post-mitotic tissues. Hence, analysis of blood mtDNAs results in the selective analysis of intermediate severity mtDNA mutations.

At the CICD, we amplified the 16.5 kb mtDNA using 59 PCR primer pairs and then sequenced the fragments by BigDye Terminator (Applied Biosystems Inc, Foster City, CA, USA) using 108 primers on an ABI 3730xl sequencer. At UCI, we resequenced the mtDNA for all cases to fill in gaps and confirm mutations. We used a similar protocol except for amplification by long-range PCR using eight primer pairs and then sequenced using 47 primers. Primer sequences are available on request.

Sequence analysis using MITOMASTER

We assembled the data for each case using Sequencher 4.7 software (Gene Codes, Ann Arbor, MI, USA) in comparison with the mtDNA reference, the revised Cambridge Reference Sequence (rCRS; NC_012920)[25] which is haplogroup H2a2a.

We evaluated the mtDNA contig using MITOMASTER, a new query system to interpret genetic variation found in mtDNA sequences.[23] For each case, we uploaded the complete mtDNA sequence into MITOMASTER to find all mtDNA variants, their gene location and evolutionary conservation. We used MITOMASTER and PhyloTree, build 5 Van Oven et al[22] to define the mtDNA

Table 1 Clinical features and sequencing results for 29 Italian patients with mitochondrial cardiomyopathy (1–21) or suspected mitochondrial disease (NonCM1-8)

Case no	Age[a]	Sex	Clinical information/follow-up	Haplogroup[b]	Total variants[c]	By variant haplogroup		By variant frequency		
						Associated	Non	Com	Rare	Novel
1	60 Years	F	HCM→DCM familial, HF, heart Tx	N1b1c	33	32	1	21	12	0
2	43 Years	F	HCM→DCM, myopathy, LBBB/living	H	14	12	2	10	2	2
3	16 Years	M	HCM/living	HV0	14	12	2	12	1	1
4	18 Years	M	Hypertrophic DCM and MR/living	H2	10	9	1	8	1	1
5	29 Years	M	SCD, HF, myopathy	U8a1a	30	29	1	20	9	1
6	61 Years	F	DCM, diabetes, hearing loss	V6	17	16	1	15	1	1
7	1 Month	F	HCM familial/died as an infant	HV4a	17	16	1	11	6	0
8	34 Years	M	HCM familial, consanguinity	H2	15	14	1	12	3	0
9	1 Year	M	HCM, severe/died	H1c3	13	13	0	12	1	0
10	13 Years	M	HCM and hearing loss	H2	10	10	0	9	1	0
11	40 Years	M	HCM familial and myopathy	H2	10	10	0	10	0	0
12	63 Years	F	HCM	H1	11	9	2	9	2	0
13	1 Year	M	DCM severe	K1c1	32	32	0	31	1	0
14	12 Years	F	DCM familial	X1a	25	24	1	21	4	0
15	26 Years	M	DCM and LHON	U5a1	25	24	1	23	1	1
16	30 Years	F	DCM and SCD of sister	H33	9	8	1	6	3	0
17	40 Years	M	DCM familial	J1c	29	29	0	26	3	0
18	45 Years	M	DCM/died at age 48 years	T2c1a	36	34	2	29	6	1
19	45 Years	F	DCM and atrial fibrillation/living	H3	7	7	0	7	0	0
20	45 Years	F	DCM familial and AV block	T2d	39	37	2	31	5	3
21	28 Years	F	HCM familial	H2	12	12	0	9	3	0
NonCM1	15 Years	M	LHON, RBBB/living, ECHO nl	U5a2a	29	26	3	26	2	1
NonCM2	40 Years	F	Myopathy familial, ECHO nl	H6a1	13	13	0	11	2	0
NonCM3	42 Years	M	LHON, ECHO nl	W1	31	31	0	30	0	1
NonCM4	18 Years	F	Ptosis familial/living, ECHO nl	H13a2b1	11	10	1	9	1	1
NonCM5	55 Years	F	Neuromuscular sx, ECHO nl	M1a1	39	37	2	37	2	0
NonCM6	26 Years	F	LHON, ECHO nl	H4a	16	13	3	13	2	1
NonCM7	62 Years	M	Lipomatosis, ECHO nl	H13a2b1	13	12	1	10	2	1
NonCM8	32 Years	M	Myopathy, ECHO nl/declined by patient	J2b1	33	32	1	30	2	1
				Total	**593**	**563**	**30**	**498**	**78**	**17**

Abbreviations: AV, atrioventricular; Com, common variants; DCM, dilated cardiomyopathy; ECHO nl, echocardiograms normal; F, female; HCM, hypertrophic cardiomyopathy; heart Tx, heart transplant; HF, heart failure; LBBB/RBBB, left/right bundle branch block; LHON, Leber's hereditary optic neuropathy; M, male; MR, mental retardation; SCD, sudden cardiac death; sx, symptoms.
[a]Age of proband at last clinical evaluation.
[b]Haplogroup was determined by MITOMASTER and PhyloTree.[22,23]
[c]The total number of mitochondrial DNA variants was obtained by comparing the sequence to NCBI reference (NC_012920). The variants were categorized 'by haplogroup' as 'associated,' variants that defined the specific haplogroup motif compared with rCRS haplogroup (H2a2a) or defined other haplogroups or as 'non,' non-motif associated variants. The variants also were categorized 'by frequency' based on allele frequencies (AF) for common variants with AF >0.5%; rare variants (rare) with AF <0.5% and novel variants (Novel) with AF=0. Case 1 was a patient from a previously reported HCM family with *MYH7* mutation.[24] The shaded rows note the six cases with potential mitochondrial DNA mutations identified in this study.

lineage (haplogroup). MITOMASTER included a database of over 3600 mtDNA sequences from NCBI GenBank. This allowed us to obtain allele frequencies for each variant. We confirmed and updated allele frequencies using analysis of Pereira et al[26] based on 5140 mtDNA GenBank sequences.

Identification of potential mtDNA mutations

Using MITOMASTER, each patient mtDNA sequence was compared with the rCRS and all deviant nucleotides identified. The array of sequence variants was used to determine the patient's haplogroup. The haplogroup-associated variants were identified and the population frequency of each nucleotide sequence variant was calculated relative to the MITOMASTER and Pereira et al[26] databases. To identify rare variants, we arbitrarily defined variants with allele frequencies <0.5% as rare and those with frequencies ≥0.5% as common, similar to cutoff used for nDNA variation.[27]

Novel and rare variants were further evaluated for their potential pathogenicity based on the gene affected, the conservation of the altered nucleotide, the predicted amino acid (AA) change if the variant occurred within a polypeptide gene and the conservation index (CI) of the altered amino acid, and the presence of heteroplasmy.[6–8,13,15–17,21,23]

Statistical analysis

We used JMP 7.0 (SAS Institute, Cary, NC, USA) for χ^2-analysis. We considered $P < 0.05$ as a significant difference.

RESULTS

mtDNA sequence variants detected by Sanger sequencing relative to rCRS

For the 29 mtDNA sequences, we observed 662 mtDNA variants compared with the rCRS (Supplementary Table). We excluded 66 highly polymorphic mtDNA variants previously described in the mtDNA control region at positions: 303–315, 522–523, 574, 16 180–16 193 and 16 519.[22,28] We excluded three known variants because of insertion or deletion of a single nucleotide (indels), as allele frequencies were not available. The indels included m.960delC in 12S rRNA (case 11) and two non-coding variants m.44insC (case 18) and m.498delC (case 13).[28] We focused on the remaining 593 total mtDNA substitution variants (Table 1) that included 270 different mtDNA variants, as multiple variants were found in more than one patient's mtDNA.

Haplogroup-associated variation accounted for most of the mtDNA variants only 17 of which were novel

In the first analysis, the 593 substitution variants were analyzed for those that were haplogroup associated. This revealed that 563 (95%) have been associated with various haplogroups, encompassing a total of 240 different variants (Table 1; Supplementary Table). This left 30 variants as 'non-haplogroup associated.' On average, there was one 'non-haplogroup-associated' variant per case, ranging from zero to three variants. In nine cases, all mtDNA variants proved to be haplogroup motifs and their mtDNAs were not analyzed further.

In the second analysis, all variants relative to the rCRS were analyzed for their frequency in the MITOMASTER database. Of the 593 mtDNA variants, 498 (84%) were present at a frequency $\geq 0.5\%$, 78 (13%) were rare ($< 0.5\%$) and 17 (3%) were novel (Table 1; Supplementary Table). On the basis of allele frequencies of the 270 different variants, 176 (65%) were common, 77 (29%) rare and 17 (6%) novel. Criteria based on allele frequency that excluded all common and rare variants was more stringent than the haplogroup criteria; 94% of the variants were excluded based on frequency, whereas 89% were excluded based on haplogroup criteria ($\chi^2 = 3.9$, df$=1$, $P=0.047$).

Six novel, non-haplogroup variants identified as potential mtDNA mutations

We excluded all haplogroup-associated variants and common or rare variants including m.513G$>$C and m.9614A$>$T, two novel variants that were haplogroup-associated variants. The conflicting classifications of these two variants were not surprising, as the mtDNA arrays that comprise MITOMASTER and PhyloTree were not completely identical. Further analysis showed that variants m.513G$>$C and m.9614A$>$T were not likely to be disease-causing mutations; m.513G$>$C was a non-coding variant and recurrent in multiple haplogroups in PhyloTree, whereas m.9614A$>$T was located in the COIII coding region but resulted in a synonymous AA change.

We then evaluated the 15 remaining novel variants as possible mutations (Table 2) based on the gene affected, sequence conservation, polypeptide amino acid substitution and conservation, and heteroplasmy.

We detected heteroplasmy for 8 of 593 variants. In all, 4 were among the 15 novel, non-haplogroup-associated variants (Table 2).

All 15 novel, non-haplogroup-associated variants were mRNA-coding variants, of which, 10 variants involved highly conserved positions (CI $\geq 67\%$). In addition, 6 of the 15 novel, non-haplogroup-associated variants were nonsynonymous variants and 9 were synonymous variants. We eliminated the nine synonymous variants and focused on the six nonsynonymous variants as a potential mutation in six different cases. Therefore, 23 of 29 cases (79%) lacked potential mtDNA substitution mutations.

The six nonsynonymous variants included m.15132T$>$C, *MT-CYB* (p.M129T) in case 2; m.15324C$>$T, *MT-CYB* (p.A193V) in case 3; m.11069A$>$G, *MT-ND4* (p.I104V) in case4; m.15222A$>$G, *MT-CYB* (p.D159G) in NonCM1; m.8954T$>$C, *MT-ATP6* (p.I143T) in NonCM6 and m.6570G$>$T, *MT-CO1* (p.A223S) in NonCM8 (Table 2; Figure 1). Of these six variants, three were heteroplasmic and three altered amino acids that are highly conserved (CI$\geq 67\%$; Table 2). Two of the six variants had overlap; that is, two variants altered conserved amino acids and also were heteroplasmic. We considered these two variants, m.15132T$>$C, *MT-CYB* (p.M129T) in case 2 and m.6570G$>$T, *MT-CO1* (p.A223S) in NonCM8, to have the greatest potential for being pathogenic mutations. For case 2 and NonCM8, a matrilineal inheritance pattern for mitochondrial disease was supported or suggested by pedigree analysis and clinical screening (Table 2, Figure 2). Molecular testing is being offered currently to available family members.

DISCUSSION

Approach to identify potential mtDNA mutations from normal variation

The goal of this study was to evaluate our MITOMASTER strategy to identify potentially pathogenic mtDNA mutations within patient mtDNA sequences. Of particular relevance was to test the effectiveness of our system for distinguishing between haplogroup variants and pathogenic mutations (Figure 3).[23,29,30]

This approach was first successfully used to identify the pathogenic mtDNA mutation that caused Leber's hereditary optic neuropathy

Table 2 Fifteen novel, non-haplogroup-associated mitochondrial DNA variants

Location/effect	Variant	Gene	AA change	CI	Heteroplasmy[a]	Haplogroup	Case no.	Matrilineal inheritance of mitochondrial disease[b]
mRNA/Nonsyn	**m.15132T$>$C**	**MT-CYB**	**p.M129T**	**0.82**	**+**	**H**	**2**	**Suggested by FH of early stroke, DM and CMP (?)**
	m.15324C$>$T	MT-CYB	p.A193V	0.46	−	HV0	3	None
	m.11069A$>$G	MT-ND4	p.I104V	0.36	−	H2	4	None
	m.15222A$>$G	MT-CYB	p.D159G	0.36	+	U5a2a	NonCM1	None
	m.8954T$>$C	MT-ATP6	p.I143T	0.77	−	H4a	NonCM6	Unknown
	m.6570G$>$T	**MT-CO1**	**p.A223S**	**0.92**	**+**	**J2b1**	**NonCM8**	**Present—maternal myopathy**
mRNA/Synon	m.12759C$>$T	MT-ND5	Silent	0.97	−	H	2	
	m.4053A$>$G	MT-ND1	Silent	0.08	−	U8a1a	5	
	m.11920C$>$T	MT-ND4	Silent	0.87	−	V6	6	
	m.7792C$>$T	MT-CO2	Silent	1.00	−	U5a1	15	
	m.9980A$>$G	MT-CO3	Silent	1.00	−	T2c1a	18	
	m.7295A$>$G	MT-CO1	Silent	0.97	−	T2d	20	
	m.9854T$>$G	MT-CO3	Silent	0.28	+	T2d	20	
	m.8266A$>$G	MT-CO2	Silent	0.67	−	H13a2b1	NonCM4	
	m.13575C$>$T	MT-ND5	Silent	0.95	−	H13a2b1	NonCM7	

Abbreviations: AA, amino acid; CI, conservation index; CMP (?), questionable diagnosis of cardiomyopathy; DM, diabetes mellitus; Nonsyn, nonsynonymous; Synon, synonymous.
[a]The presence (+) or (−) absence of heteroplasmy detected by Sanger sequencing are provided.
[b]Inheritance based on the results of pedigree analysis and clinical screening by echocardiography of available family members; pedigrees of Case 2 and NonCM8 are shown in Figure 2.
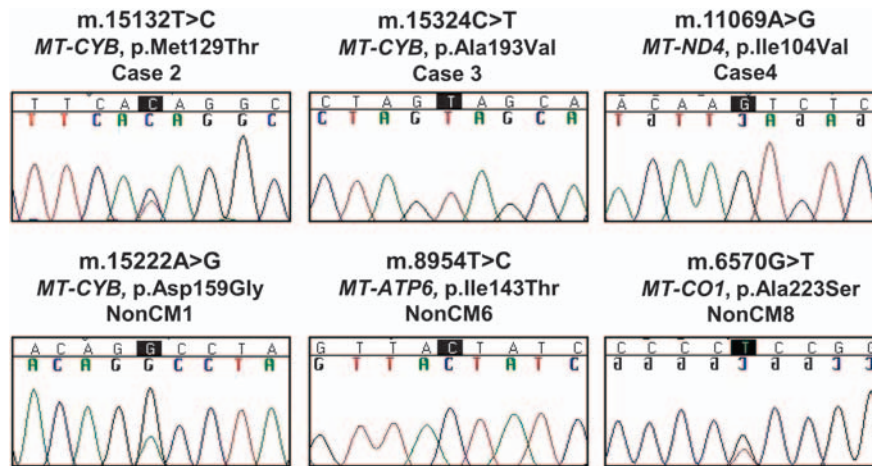The variants with the greatest pathogenic potential are in bold.

Figure 1 Six novel, non-haplogroup variants as potential pathogenic mtDNA mutations. Shown are chromatograms for six potential mtDNA mutations m.15132T>C, *MT-CYB* (p.M129T) in case 2; m.15324C>T, *MT-CYB* (p.A193V) in case 3; m.11069A>G, *MT-ND4* (p.I104V) in case4; m.15222A>G, *MT-CYB* (p.D159G) in NonCM1; m.8954T>C in *MT-ATP6* (p.I143T) in NonCM6; m.6570G>T, *MT-CO1* (p.A223S) in NonCM8. Heteroplasmy was detected in three, that is, case 2, NonCM1 and NonCM8.
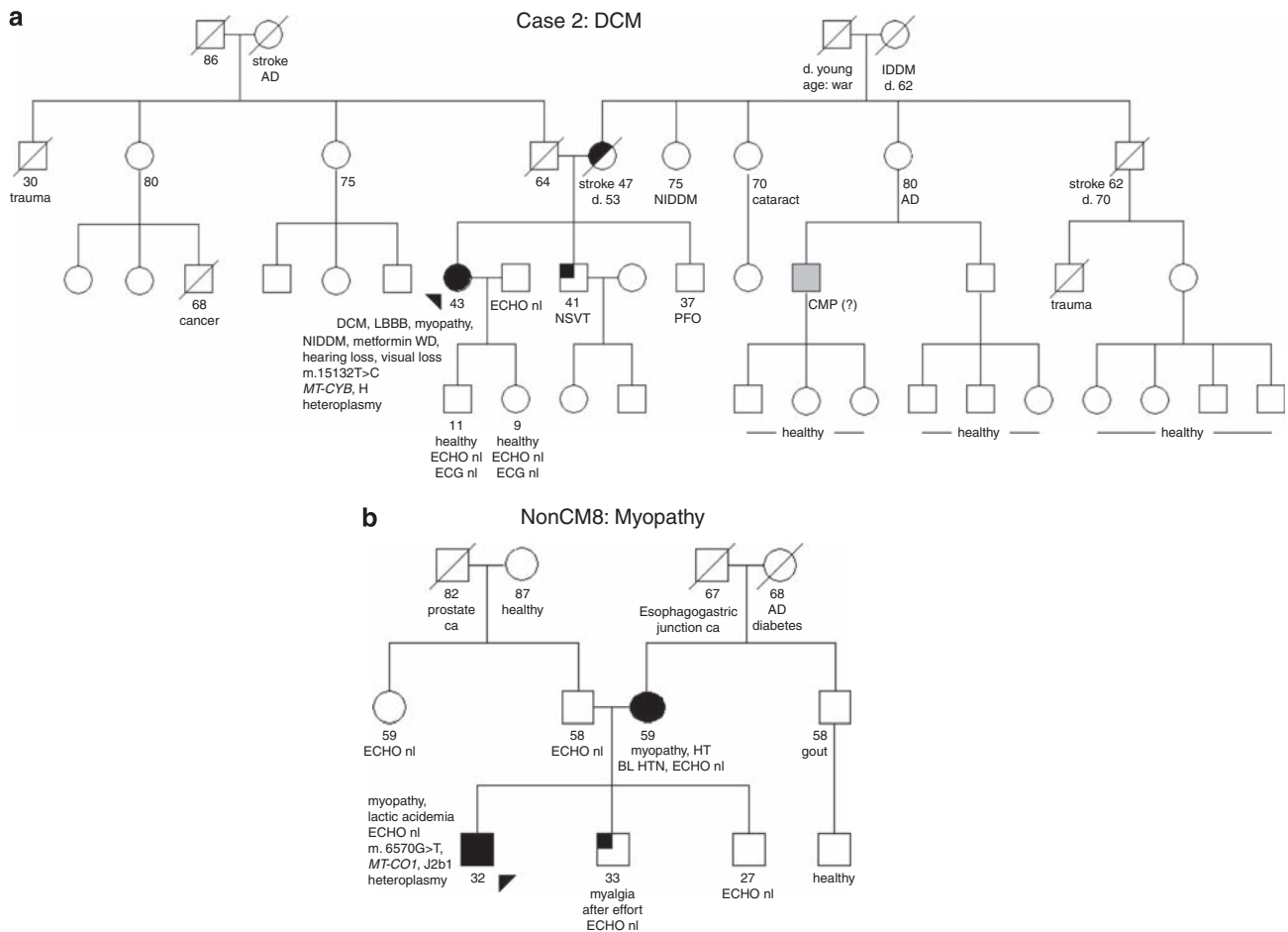

Figure 2 Pedigree and clinical features for two cases with novel, potential mtDNA mutations: case 2 (m.15132T>C, *MT-CYB*) and NonCM8 (m.6570G>T, *MT-CO1*). Pedigree analysis and clinical screening by echocardiography suggested (case 2) or supported (NonCM8) matrilineal inheritance of mitochondrial disease. Circles represent females and squares represent males; solid shapes are affected individuals; an arrow indicates the proband. Number is the age in years at last clinical evaluation or age of death. (**a**) Case 2: dilated cardiomyopathy (DCM). Family history included maternal relatives with early stroke (<55 years), diabetes and questionable cardiomyopathy. (**b**) NonCM8: myopathy. Family history included the mother with myopathy and maternal grandmother with diabetes. Abbreviations: AD, Alzheimer's disease; BL HTN, borderline hypertension; ca, cancer; d., died; CMP (?), questionable diagnosis of cardiomyopathy by history; ECG nl, electrocardiogram results normal; ECHO nl, echocardiography results normal; HT, hypertriglyceridemia; LBBB, left bundle branch block; NIDDM, non-insulin dependent diabetes mellitus; NSVT, nonsustained supraventricular tachycardia; PFO, patent foramen ovale.
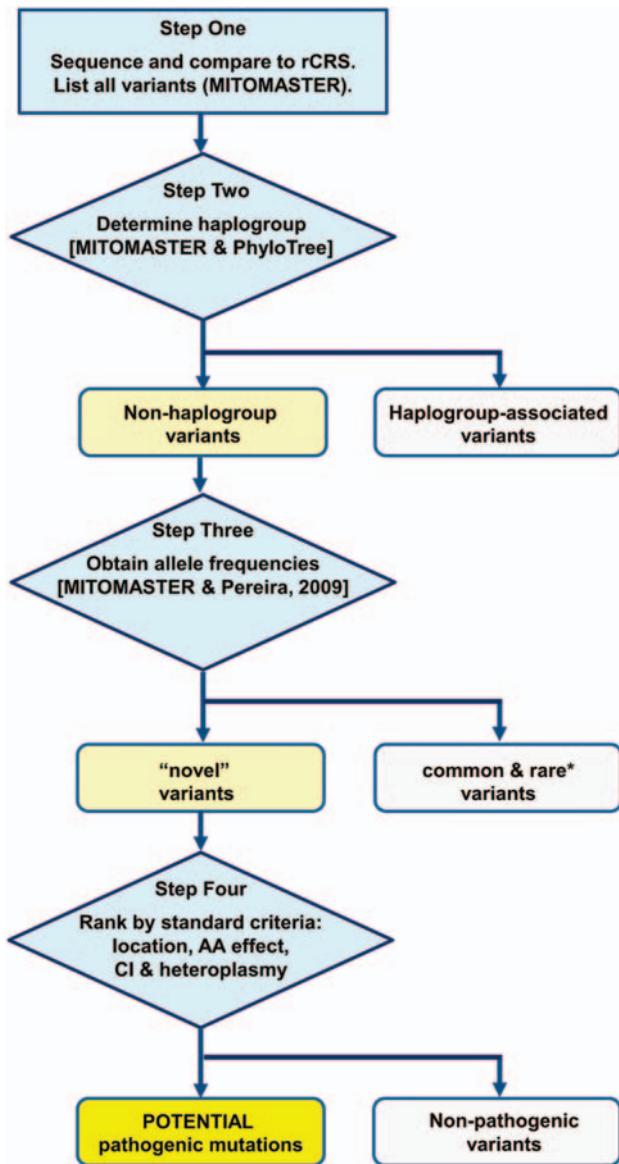
Figure 3 Approach to identify potential mtDNA mutations. Four major steps to facilitate the selection of potential mtDNA mutations. *Rare variants require additional analysis before exclusion.

associated OXPHOS complex I defect.[32] The results confirmed m.14459G>A as a mutation, and a general approach to find potential mtDNA mutations was described.

In this study, we applied this approach to a large number of mtDNA sequences from patients with cardiomyopathy or another potential mitochondrial disease phenotype. These patient sequences were compared with thousands of mtDNA sequences using our automated analysis system, MITOMASTER. The approach used two additional factors, haplogroup and allele frequencies, along with standard criteria to evaluate the long list of mtDNA variants detected in molecular studies on cardiomyopathy.[6–8,15–17] To access large numbers of reference mtDNAs, we used online resources such as GenBank. Using this system, we were able to eliminate 98% of the 270 different variants and excluded 79% of the cases (23 of 29).

Analysis of published mtDNA variants is another application using this system. We tested the approach in identifying the confirmed m.14459G>A mutation among the 40 variants of the LHON patient.[31] Using the approach, the majority of the variants were excluded as haplogroup-associated variants including D1 haplogroup variants. The m.14459G>A mutation was the only candidate mutation left after analysis of the non-haplogroup variants using allele frequencies and standard criteria. We conducted example runs with published mtDNA variants not yet confirmed and found evidence for erroneous classifications of haplogroup-associated variants as mutations. We believe that this approach may be used to identify potential mutations among the >200 mtDNA variants reported in cardiomyopathy patients;[6–8,15–17] research studies may then be concentrated for validation of these selected variants.

We emphasize that the approach in this study was presented to facilitate the selection of potential mtDNA mutations. Functional analysis, clinical evaluations, pedigree analysis, family screening and follow-up are critical in the assessment of potential mtDNA mutations.[1,2,5] Sub-haplogroup motifs particularly at the ends of the haplogroup tree need careful consideration, as they may be relatively new and rare in the population.[21] It is possible that these nucleotides contributed to disease as a primary defect or in combination with a nDNA mutation. For example, rare variants found in the N1b1c sub-haplogroup for case 1 may have influenced the progression of *MYH7*-hypertrophic cardiomyopathy,[24] unpublished results. Thus, without additional analyses, the contribution of rare variants or haplogroup effects to disease may be missed. Validation of this approach will require transmitochondrial cybrid studies of the potential mutations and sets excluded in this approach (for example, sub-haplogroup motifs).

Potential mtDNA mutations for mitochondrial disease

By the end of our analysis, we had narrowed the potentially pathogenic mtDNA mutations to six. These included one each in *MT-CO1* (m.6570G>T), *MT-ATP6* (8954T>C) and *MT-ND4* (m.11069A>G) and three in *MT-CYB* (m.15132T>C, m.15222A>G and m.15324C>T). These six potential mutations were not present among the confirmed mutations for mitochondrial disease in MITOMAP or in previous studies on cardiomyopathies.[6–8,15–17,28] Of the six, m.15132T>C (p.M129T) and m.6570G>T (p.A223S) have the greatest potential of being pathogenic based on high evolutionary conservation (CI=0.82 and 0.92, respectively) and the detection of heteroplasmy. In addition, clinical evidence for matrilineal inheritance of mitochondrial disease was present or suggested in the family of case 2 and NonCM8 (Table 2; Figure 2).

The final candidate mutations included two heteroplasmic *CytB* variants: m.15132T>C for case 2 (haplogroup H) with hypertrophic

(LHON) and dystonia in a Hispanic family.[29,31] Complete mtDNA sequencing of an affected family member revealed 40 variants compared with reference sequence (step 1). Phylogenetic analysis revealed that the patient's mtDNA belonged to native American haplogroup D, permitting exclusion of haplogroup-associated variants including the previously unreported variant m.2092C>T (step 2). Additional variants were excluded by their being reported in the literature at 'significant polymorphism frequencies' (step 3). One novel, non-haplogroup variant remained, m.14459G>A in NADH dehydrogenase subunit 6 (*MT-ND6*). This variant altered a moderately conserved AA in the ND6 polypeptide. Detection of heteroplasmy provided further support of m.14459G>A as a potential mutation (step 4).

Functional evidence of pathogenicity came later from transmitochondrial cybrid studies of m.14459G>A, which revealed an

DCM and m.15222A>G for NonCM1 (haplogroup U5a2a) with LHON and right bundle branch block. A third likely pathogenic mutation was a homoplasmic *MT-ATP6* variant m.8954T>C (haplogroup H4a) for NonCM6 with LHON based also on its high level of conservation (CI=0.77).[21]

Three of the six potential mutations have been proposed previously to be pathogenic, namely, heteroplasmic m.6570G>T variant in *MT-CO1* (haplogroup J2b1; CI=0.92) for NonCM8 with myopathy; homoplasmic m.11069A>G in *MT-ND4* (haplogroup H2; CI=0.36) for case 4 with HCM, mild DCM and mental retardation; and homoplasmic m.15324C>T in *MT-CYB* (haplogroup HV0; CI=0.36) for case 3 with HCM. The m.6570G>T variant was observed as a novel germline variant in a patient with sporadic parathyroid adenoma.[33] The 11069 and 15324 variants have been reported in prostate cancer tissue,[34] the specific haplogroups of these patients were not provided. The m.11069A>G also was found as one of 13 mtDNA variants in a patient with Parkinson's disease on a haplogroup H2 background,[35] the same lineage as our patient. Taken together with low level of conservation, it is possible that m.11069A>G and m.15324C>T, two 'novel' homoplasmic variants, are haplogroup-associated variants and not mutations.

mtDNA sequence databases: the need for haplogroup-specific control data

This analysis clearly demonstrates the critical need for large arrays of good quality mtDNA sequences from normal individuals that encompass the full array of mtDNA lineages from around the world. Furthermore, it demonstrates that the analysis of individual mtDNA sequences from patients cannot be managed by hand, but must be developed using automated systems such as MITOMASTER.[23]

Currently, the total extent of normal mtDNA variation within the multitude of haplogroups from the various continental populations remains unknown. Therefore, mtDNAs of several thousand will be required from each continent to capture variants with frequencies in the range of <0.5% and <0.1%, similar to cutoffs used for nDNA variation.[36]

In summary, the approach demonstrated in this study allowed us to find potential mutations in patients with mitochondrial cardiomyopathy. We hope this approach will become more powerful as we obtain more data on the depth of normal mitochondrial variation throughout the world. As we enter the new era of low cost, high-throughput genome sequencing, large scale sequencing efforts will hasten these efforts. This will help us understand the role of the highly variable mitochondrial genome in human disease.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

1 Wallace DC, Brown MD, Lott MT: Mitochondrial DNA variation in human evolution and disease. *Gene* 1999; **238**: 211–230.
2 Marin-Garcia J, Goldenthal MJ: Cardiomyopathy and abnormal mitochondrial function. *Cardiovasc Res* 1994; **28**: 456–463.
3 Anderson S, Bankier AT, Barrell BG *et al*: Sequence and organization of the human mitochondrial genome. *Nature* 1981; **290**: 457–465.
4 Wallace DC: A mitochondrial paradigm of metabolic and degenerative diseases, aging, and cancer: a dawn for evolutionary medicine. *Ann Rev Genet* 2005; **39**: 359–407.
5 Brega A, Narula J, Arbustini E: Functional, structural, and genetic mitochondrial abnormalities in myocardial diseases. *J Nucl Cardiol* 2001; **8**: 89–97.
6 Obayashi T, Hattori K, Sugiyama S *et al*: Point mutations in mitochondrial DNA in patients with hypertrophic cardiomyopathy. *Am Heart J* 1992; **124**: 1263–1269.
7 Li YY, Maisch B, Rose ML, Hengstenberg C: Point mutations in mitochondrial DNA of patients with dilated cardiomyopathy. *J Mol Cell Cardiol* 1997; **29**: 2699–2709.
8 Arbustini E, Diegoli M, Fasani R *et al*: Mitochondrial DNA mutations and mitochondrial abnormalities in dilated cardiomyopathy. *Am J Pathol* 1998; **153**: 1501–1510.
9 Anan R, Nakagawa M, Miyata M *et al*: Cardiac involvement in mitochondrial diseases. A study on 17 patients with documented mitochondrial DNA defects. *Circulation* 1995; **91**: 955–961.
10 Vilarinho L, Santorelli FM, Rosas MJ, Tavares C, Melo-Pires M, DiMauro S: The mitochondrial A3243G mutation presenting as severe cardiomyopathy. *J Med Genet* 1997; **34**: 607–609.
11 Goto Y, Nonaka I, Horai S: A mutation in the tRNA(Leu)(UUR) gene associated with the MELAS subgroup of mitochondrial encephalomyopathies. *Nature* 1990; **348**: 651–653.
12 Manouvrier S, Rötig A, Hannebique G *et al*: Point mutation of the mitochondrial tRNA(Leu) gene (A3243G) in maternally inherited hypertrophic cardiomyopathy, diabetes mellitus, renal failure, and sensorineural deafness. *J Med Genet* 1995; **32**: 654–656.
13 Zeviani M, Gellera C, Antozzi C *et al*: Maternally inherited myopathy and cardiomyopathy: association with mutation in mitochondrial DNA tRNA(Leu)(UUR). *Lancet* 1991; **338**: 143–147.
14 Nan DN, Fernández-Ayala M, Infante J, Matorras P, González-Macías J: Progressive cardiomyopathy as manifestation of mitochondrial disease. *Postgrad Med J* 2002; **78**: 298–299.
15 Marin-Garcia J, Goldenthal MJ, Ananthakrishnan R, Pierpont ME: The complete sequence of mtDNA genes in idiopathic dilated cardiomyopathy shows novel missense and tRNA mutations. *J Card Fail* 2000; **6**: 321–329.
16 Ruppert V, Nolte D, Aschenbrenner T, Pankuweit S, Funck R, Maisch B: Novel point mutations in the mitochondrial DNA detected in patients with dilated cardiomyopathy by screening the whole mitochondrial genome. *Biochem Biophys Res Commun* 2004; **318**: 535–543.
17 Schrijver I, Pique LM, Traynis I, Scharfe C, Sehnert AJ: Mitochondrial DNA analysis by multiplex denaturing high-performance liquid chromatography and selective sequencing in pediatric patients with cardiomyopathy. *Genet Med* 2009; **11**: 118–126.
18 Johnson MJ, Wallace DC, Ferris SD, Rattazzi MC, Cavalli-Sforza LL: Radiation of human mitochondria DNA types analyzed by restriction endonuclease cleavage patterns. *J Mol Evol* 1983; **19**: 255–271.
19 Merriwether DA, Clark AG, Ballinger SW *et al*: The structure of human mitochondrial DNA variation. *J Mol Evol* 1991; **33**: 543–555.
20 Mishmar D, Ruiz-Pesini E, Golik P *et al*: Natural selection shaped regional mtDNA variation in humans. *Proc Natl Acad Sci USA* 2003; **100**: 171–176.
21 Ruiz-Pesini E, Mishmar D, Brandon M, Procaccio V, Wallace DC: Effects of purifying and adaptive selection on regional variation in human mtDNA. *Science* 2004; **303**: 223–226.
22 van Oven M, Kayser M: Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Hum Mutat* 2009; **30**: E386–E394. http://www.phylotree.org.
23 Brandon MC, Ruiz-Pesini E, Mishmar D *et al*: MITOMASTER: a bioinformatics tool for the analysis of mitochondrial DNA sequences. *Hum Mutat* 2009; **30**: 1–6.
24 Arbustini E, Fasani R, Morbini P *et al*: Coexistence of mitochondrial DNA and beta myosin heavy chain mutations in hypertrophic cardiomyopathy with late congestive heart failure. *Heart* 1998; **80**: 548–558.
25 Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N: Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 1999; **23**: 147.
26 Pereira L, Freitas F, Fernandes V *et al*: The diversity present in 5140 human mitochondrial genomes. *Am J Hum Genet* 2009; **84**: 628–640.
27 Manolio TA, Collins FS, Cox NJ *et al*: Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
28 MITOMAP. website: http://www.mitomap.org date accessed November 2009.
29 Torroni A, Wallace DC: Mitochondrial DNA variation in human populations and implications for detection of mtDNA mitochondrial DNA mutations of pathological significance. *J Bioenerg Biomembr* 1994; **26**: 261–271.
30 Ruiz-Pesini E, Lott MT, Procaccio V *et al*: An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic Acids Res* 2007; **35**: D823–D828.
31 Jun AS, Brown MD, Wallace DC: A mitochondrial DNA mutation at nucleotide pair 14459 of the NADH dehydrogenase subunit 6 gene associated with maternally inherited Leber hereditary optic neuropathy and dystonia. *Proc Natl Acad Sci USA* 1994; **91**: 6206–6210.
32 Jun AS, Trounce IA, Brown MD, Shoffner JM, Wallace DC: Use of transmitochondrial cybrids to assign a complex I defect to the mitochondrial DNA-encoded NADH dehydrogenase subunit 6 gene mutation at nucleotide pair 14459 that causes Leber hereditary optic neuropathy and dystonia. *Mol Cell Biol* 1996; **16**: 771–777.

33 Costa-Guda J, Tokura T, Roth SI, Arnold A: Mitochondrial DNA mutations in oxyphilic and chief cell parathyroid adenomas. *BMC Endocr Disord* 2007; **7**: 8.

34 Parr RL, Dakubo GD, Crandall KA *et al*: Somatic mitochondrial DNA mutations in prostate cancer and normal appearing adjacent glands in comparison to age-matched prostate samples without malignant histology. *J Mol Diagn* 2006; **8**: 312–319.

35 Vives-Bauza C, Andreu AL, Manfredi G *et al*: Sequence analysis of the entire mitochondrial genome in Parkinson's disease. *Biochem Biophys Res Comm* 2002; **290**: 1593–1601.

36 Ionita-Laza I, Lange C, M Laird N: Estimating the number of unseen variants in the human genome. *Proc Natl Acad Sci USA* 2009; **106**: 5008–5013.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)