RESEARCH ARTICLE

Application of Phylogenetic Analysis in the Study of Mitochondrial Genetic Diseases

Ertan Kurt, Rizal Masdari
University of Life Science Research, Vancouver

Abstract: Mitochondrial genetic diseases are receiving increasing attention. However, the common case-control methods are susceptible to genetic background, population stratification, data quality and other factors that may lead to false positive results. The field of genetic diseases has not been paid enough attention. Therefore, this paper reviews this method and promotes the extensive and in-depth use of mitochondrial genetic diseases.

Keywords: Mitochondrial; phylogenetic; haplotype; data quality

Mitochondrial genetic diseases are diseases caused by MITOCHON DRIAL DNA (mtDNA) deficiencies (including MITOCHON DRIAL DNA deletion, insertion and point mutation) that lead to MITOCHON DRIAL dysfunction. Mitochondria are organelles that provide energy within cells and regulate programmed cell death. Dysfunction of the mitochondria can lead to a series of disorders, such as dyssynthesis of ATP and insufficient energy production. In addition to LHON (Lieber's hereditary optic atrophy), Leigh disease (subacute necrotizing encephalomyelopathy), MELAS (mitochondrial encephalomyopathy with lactoxemia and stroke-like attack) and MERF (muscle), which have been well described clinically In addition to many clinical syndromes such as clonic epilepsy with RRF, patients with mitochondrial dysfunction may exhibit a variety of other systemic abnormalities, such as deafness, diabetes, heart disease, tumor[1-4] Holt *et al*.[5] for the first time in 1988, mitochondrial DNA deletions were found in patients with mitochondrial diseases, suggesting that mtDNA mutations may be severe. The cause of the disease.

With the deepening of research on mitochondrial genetic diseases, more and more people have come to understand the role of mtDNA in diseases. However, there are also some problems in this field, mainly as follows: (1) ignoring the ancestral variation of mtDNA itself, that is, maternal genetic background, is easy to mutate on mtDNA. Isolated views often lead to false positive results or inaccurate inferences or even incorrect conclusions. Current studies of mitochondrial genetic diseases usually use case-control methods of association analysis, i. e. comparing mtDNA mutations between patients and normal controls to identify possible mutations. However, because the mitochondrial genome is a non-recombinant molecule, all mutation sites are tightly linked. A simple case-control analysis without taking into account the entire variation pattern of the haplotype to which the mutation belongs may treat other mutation sites closely linked to the pathogenic mutation as pathogenic mutations, thus obtaining false positive results. Results. For example, many mtDNA mutations previously thought to be associated with disease or longevity are actually characteristic mutations in some lineages[6–8]. Statistically significant differences exist between groups, and this mutation is often thought to be associated with disease occurrence, but this method is very vulnerable to the effect of population stratification resulting in false positive results. Population stratification is more complicated by recombinant, wholly linked molecules. Different subpopulations within a population may belong to different mtDNA lineages, carrying different characteristic mutation sites, and even patients from the same region and normal controls may belong to different mtDNA lineages. For normal controls, the characteristic variation in the accumulation of mtDNA in the patient's lineage itself

doi: 10.18063/gds.v2i1.

may be mistaken for a pathogenic mutation. For this reason, many mtDNA mutations associated with disease obtained from case-control methods are often difficult to replicate in studies of different populations. In recent years, more and more mtDNA sequences have been obtained, but unfortunately, due to the lack of strict quality control measures, the reported mtDNA sequence data in this field often have one or more problems and errors, leading to serious data quality problems in mitochondrial genetic diseases. This also has some doubts about its findings or inferences[11–13].

In view of the above problems, phylogenetic analysis should be recommended: under the guidance of reconstructed mtDNA phylogenetic relationships, the maternal genetic structure of the subjects should be assessed first, and then the combination of phylogenetic knowledge can help to identify characteristic mutation sites and potential pathogenic mutations, thus avoiding erroneous conclusions. This paper reviews the application of phylogenetic analysis in mitochondrial genetic diseases, hoping to be helpful to the study of mitochondrial genetic diseases in China.

1. The concept of phylogenetic relationship of 1mtDNA

The phylogeny is relative to ontogeny. It refers to the formation and development of a certain group. Each group has its own history of development, so what is the evolutionary relationship between different groups? To explore this problem is to study the phylogeny of different groups. The origin and evolutionary relationship of population, that is, reconstructing the phylogenetic relationship of the world population, is a very important part of evolutionary biology, and has attracted more and more attention.

Human mtDNA is a ring-closed molecule. Because of its strict maternal inheritance and no recombination, a mtDNA genome corresponds to haplotype. In a maternal-radiated mtDNA haplotype lineage, mutations in the mitochondrial genome originate purely from the sequential accumulation of new mutations. In addition, all mutations can be faithfully and continuously inherited along the maternal line, so there is theoretically a temporal sequence of all mutations accumulated in a mtDNA lineage. After the population spread to different areas, the mtDNA lineages in the population still accumulate and mutate continuously, resulting in the generation-specific derivative lineages. Since there is no recombination of mtDNA, although the lineages are still accumulating and mutating gradually in their own evolution process, the lineages that already exist among the lineages are still existing. The phylogenetic relationship is not affected by the subsequent demography of the population in which it is located. Therefore, there is a theoretical evolutionary link between the mtDNA lineages in modern populations. It is possible to reconstruct the phylogenetic relationships of matrilineal lineages in a region or even in the world population. The phylogenetic tree (hereinafter referred to as "phylogenetic tree") is the most possible true reflection of the evolutionary relationships and evolutionary processes of various lineages existing in a population[14]. Group history can also help people understand and study diseases related to mtDNA mutation.

2. Application of phylogenetic analysis in mitochondrial genetic diseases

The ultimate goal of studying the phylogenetic relationship of mtDNA is not only to reconstruct the phylogenetic tree, but also to study the population relationship, population dynamics and the correlation between mutation and disease in a certain area through the mtDNA phylogenetic tree[15]. The mitochondrial phylogenetic tree has been widely connected in the field of human population genetics. Many studies on mitochondrial diseases do not or seldom refer to and use phylogenetic analysis as an effective method, which has two adverse consequences: one is the lack of quality detection of mtDNA mutation data. Secondly, the absence of phylogenetic relationships as a background for reference analysis makes it easy to look at mutations in mtDNA in isolation, resulting in false positive results or failure to reveal true pathogenic mutations.

2.1 MtDNA family tree in detecting data errors

There is every reason to doubt the authenticity and credibility of the results of this study for a batch of data with quality problems, no matter how fancy the subsequent analysis methods and how perfect the logic of the argument are. There are many data quality problems in the research area[11,12], which directly leads to many research results in the field can not be duplicated and many views are controversial and so on. One of the ideas they follow is the analysis of phy-

logenetics. Based on this idea, the following is a summary of our previous work.

When we analyze the sequence data of mtDNA, we should first take the revised Cambridge reference sequence (rCRS)[19] as a reference, and output the mutations in the sequence as mutation sites. Then we should inspect the quality of this batch of data, first we should observe whether there is a large number of sparse in the sequence. Previous large numbers of human mtDNA sequence information indicated that the frequency of transition in mtDNA was much higher than that of tran svers ion. Base tran svers ion was relatively rare compared with tran svers ion, and the mutation to G was the rarest[20–25]. More rare transmutations were observed in the 560 sequences reported by Herrnstadt *et al*.[26], including more G transmutations, and the larger of these mutations. Some of them were proved to be caused by human or sequence errors after re-experiment by the original author[27]. Secondly, phylogenetic analysis of the sequences was carried out. Individuals were divided into haplotype groups, that is, under the guidance of lineage trees, all individuals were divided into corresponding groups by comparing the same or similar variations among sequences. In theory, each copy of mtDNA has its place in the mtDNA lineage tree worldwide. At present, the backbone and branches of the mitochondrial lineage tree worldwide (especially in East Asia and Europe) have been basically clear and cannot be classified as an individual due to the emergence of a large number of complete mitochondrial sequences. Systemic status is rare. After determining the individual's phylogenetic status, any mutation that contradicts the motif of the haplotype group must be confirmed by another experiment if a rare mutation is present at different sites. Multiple occurrences in the inherited background (i. e. haplotype groups) mean that the mutation may be questionable. This is because mtDNA is maternal inheritance, lack of recombination, and mutations in evolutionary events are generally considered to follow the principle of minimalism, so rare mutations may be wrong in multiple lineages. It is absolutely necessary to obtain high-quality data to clearly reveal potential pathogenic mutations, to re-check and even re-sequence the gel map for observed rare mutations or any other suspected mutations[28]. Conflicting with the variation pattern of another fragment, there may be an artificial recombination in which two or more different haplotypes of mtDNA fragments are mixed in a sequence. In this case, the experiment (including PCR amplification and DNA sequencing) will be re-validated. The incompatibility between mutations of an individual may be caused by frequent mutations or human errors, and judgment of the problem needs to be re-carried out. Finally, for the system tree constructed on the basis of the obtained data, any sites with frequent mutation events should be re-sequenced even in the individual rare or private mutation at the top. To re test to confirm[8].

Yao *et al*.[29] examined a batch of mtDNA data published by Silva *et al*.[30] on Am J Hum Genet using the above method and found many problems and errors. At the same time, six possible causes of these problems and errors were summarized in order to prevent the recurrence of similar problems in future research work. The suggestion of Ao *et al*.[29] re-experimentally analyzed the data, confirming and correcting the problems and errors pointed out by Yao *et al*.[29] in the data. This example fully demonstrates the practicability and reliability of phylogenetic analysis for data quality detection.

2.2 MtDNA genealogy and genetic diseases

Previous studies have found that the vast majority of mutations in mtDNA are neutral[31,32], and only a few are harmful. Because of their different harmfulness and other related factors, the distribution of these mutations in the lineage trees we acquire will show a certain pattern: at the base of the mitochondrial lineage tree. Ancient and subancient mutations can be considered at least selective neutral because they have undergone tens of thousands of years of selective stress. Severe deleterious mutations can cause multiple system dysfunction, leading to very serious disease in patients. Individuals with such mutations are often unable to survive into adulthood. Thus this mutation can only exist in the form of rare mutations in different individuals and is quickly eradicated, not retained in lineages (e. g. 3243 mutation)[33]. Medium or late stages of the mutation exhibit pathogenicity and have little effect on the growth of the individual. The weaker selectivity of the mutation makes it possible for the mutation to exist in a limited number of generations and may therefore be polymorphic in the population[31,33]; and in some special cases, the mutation may occur in the population. There is a certain degree of diffusion.

Based on the above inference, we can theoretically make a preliminary judgment on the pathogenicity of the re-

ported mtDNA mutation. If a specific coding region mutation is found in the mtDNA of a patient with a maternal genetic disease, the mutation can be preliminarily classified as a class by detecting a normal individual belonging to the same haplotype group. For example, Kong *et al*.[35] found a T12338C mutation in Chinese studies that led to the loss of the starting codon of the mtDNA-ND5 gene (i. e. the translation of the starting amino acid from methionine to threonine), so it looked like a serious pathogenic mutation. According to a comprehensive analysis of 3 000 individuals from all over the country, the mutation T12338C caused the loss of the initial codon of the mtDNA ND5 gene, but as one of the characteristic mutations in the F2 group, the mutation occurred about 42, 000 years ago and was widely distributed among the general population of China (although the frequency was low), and it is now combined with No Association of F2 with mtDNA genetic diseases has been reported. The above evidence suggests that T12338C is unlikely to be a pathogenic mutation.

With detailed phylogenetic knowledge as the background, false positive results can be effectively avoided by isolating mutation sites. Because of the frequent migration and communication among modern populations, the genetic structure within populations is often very complex, and even populations living in the same region may have different genetic backgrounds. Composition of subpopulations, together with the unique genetic characteristics of mtDNA, i. e. the distinct mutations carried by different lineages and the strict non-recombination between them, makes it very inconvenient for the case-control method commonly used in genetic disease research. For example, a family with mitochondrial genetic diseases, all of which are maternal inheritance The patient's mtDNA belongs to the same haplotype, and the haplotype lineage has its own mutation. In this study, normal people from the same area were selected as controls, and it is likely that these controls belong to other haplotype groups, so there is no mutation site shared with the patient's haplotype. In this case, the patient's specific mutation site For example, Tzen *et al*.[36] found a mutation of A14693G in the mtDNA of patients with MELAS and thought it was associated with the disease. However, by analyzing the reconstructed East Asian mtDNA lineage tree, Bandelt *et al*.[37] found that the mutation was one of the characteristic sites of group Y. However, the population is widespread among the normal population in East Asia, suggesting that the mutation is more likely to be normal.

In addition, detailed phylogenetic knowledge can also be used as a background to avoid false positive or false negative results in association analysis resulting from population stratification. For example, the mitochondrial haplotype groups B, F, R9b, M7b are more abundant in southern China, while the haplotype groups A, C, D, G, M8a, Y, Z are more abundant in northern China, but if the southern population studied is moved from the northern China, the hidden stratification interference can not be completely eliminated. In the course of the study, if one does not understand the possible genetic structure differences caused by this migration event and selects such a population as a control group, one can draw some false positive conclusions about disease-related haplotype groups. Understanding the actual maternal genetic structure of populations on various continents and understanding the population relationship and population dynamics in a region can effectively assess or avoid the bias caused by population stratification[8].

Thus, any hasty conclusions or conjectures based solely on rough case-control or statistical tests when studying disease-related mtDNA mutations, without judging the molecular basis of the mutation through knowledge of mtDNA phylogeny and validating its pathogenicity through population data, can be drawn. Kong *et al*.[8,18] constructed a comprehensive and fine East Asian mtDNA phylogenetic tree by analyzing more than 1000 mtDNA whole genome sequences. The detailed phylogenetic tree and the author's suggestions on how to identify pathogenic mutations or maternal genetic backgrounds are all very good for mitochondrial inheritance. It provides reference and help for disease research.

3. Summary

The whole genome of mtDNA has developed rapidly in the past few years, and a great deal of research results have been made. The whole genome sequence of mtDNA provides the highest molecular resolution. With the increase and improvement of the quantity and quality of the whole sequence data, the mtDNA lineage tree of the world population will become more complete. Reliable phylogenetic trees can help people understand the evolution of mtDNA mutations and thus help to study genetic diseases more effectively.

References

1. prezanttr, agapianjv, bo hlmanmc, *et al*. mitoch on dria lri boso malr namuta tionas sociate dwith bo than tibiotic-inducedand non syndromicdeafness. natgenet, 1993, 4 (3): 289 - 294.
2. Petros JA, Baum ann AK, Ruiz- Pesini E, *et al*. Mt DNA mutation sincreasetumorigeni city inprostate cancer. ProcNatlAcadSciUSA, 2005, 102(3):719—724.
3. PoultonJ, LuanJ, MacaulayV, *et al*. Type2dia be tesis ass ociate dwith acomm onmitochondrialvariant:Evidencefromapopula-tion-base dcase-controlstudy. HumMolGenet, 2002, 11(13):1581—1583.
4. TanakaM, InoH, OhnoK, *et al*. Mitochondrialmutationinfatalinfantilecardiomyopathy. Lancet, 1990, 336(8728):1452.
5. HoltIJ, HardingAE, Morgan-HughesJA. DeletionsofmusclemitochondrialDNAinpatientswithmitochondrialmyopathies. Nature, 1988, 331(6158):717—719.
6. YaoYG, SalasA, BraviCM, *et al*. AreappraisalofcompletemtDNAvariationinEastAsianfamilieswithhearingimpairment. HumGenet, 2006, 119(5):505—515.
7. YaoYG, KongQP, ZhangYP. Mitochondrial DNA 5178 Apolymorphismandlongevity. HumGenet, 2002, 111(4-5):462—463.
8. KongQP, BandeltHJ, SunC, *et al*. UpdatingtheEastAsianmtDNAphylogeny:Aprerequisitefortheidentificationofpathogenicmutations. HumMolGenet, 2006, 15(13):2076—2086.
9. CardonLR, PalmerLJ. Populationstratificationandspuriousallelic association. Lancet, 2003, 361(9357):598—604.
10. ReichDE, GoldsteinDB. Detectingassociationinacase-controlstudywhilecorrectingforpopulationstratification. GenetEpide-miol, 2001, 20(1):4—16.
11. BandeltHJ, AchilliA, KongQP, *et al*. Low"penetrance"ofphylogeneticknowledgeinmitochondrialdiseasestudies. BiochemBio-physResCommun, 2005, 333(1):122—130
12. BandeltHJ, KongQP, ParsonW, *et al*. Moreevidencefornon-maternalinheritanceofmitochondrialDNA?JMedGenet, 2005, 42(12):957—960.
13. BandeltHJ, LahermoP, RichardsM, *et al*. Det ectin gerrorsinmt DNA data by phy logene ticanalysis. Int JL ega lMed, 2001, 115(2):64—69.
14 Kong Qingpeng. Mitochondrial DNA phylogenetic genomics of East Asian population. PhD dissertation. Beijing: Graduate School of Chinese Academy of Sciences, 2005.
15. PosadaD, CrandallKA. Intraspecificgenegenealogies:Treesgraftingintonetworks. TrendsEcolEvol, 2001, 16(1):37—45.
16. YaoYG, KongQP, WangCY, *et al*. Differentmatrilinealcontributionstogeneticstructureofethnicgroupsinthesilkroadregion inchina. MolBiolEvol, 2004, 21(12):2265—2280.
17. SunC, KongQP, PalanichamyMG, *et al*. ThedazzlingarrayofbasalbranchesinthemtDNAmacrohaplogroupMfromIndiaasin-ferredfromcompletegenomes. MolBiolEvol, 2006, 23(3):683—690.
18. KongQP, YaoYG, SunC, *et al*. PhylogenyofeastAsianmitochondrialDNAlineagesinferredfromcompletesequences. AmJ HumGenet, 2003, 73(3):671—676.
19. AndrewsRM, KubackaI, ChinneryPF, *et al*. ReanalysisandrevisionoftheCambridgereferencesequenceforhumanmitochondrial DNA. NatGenet, 1999, 23(2): 147.
20. Maca-MeyerN, GonzalezAM, LarrugaJM, *et al*. Majorgenomicmitochondriallineagesdelineateearlyhumanexpansions. BMC Genet, 2001, 2:13.
21. KivisildT, TolkHV, ParikJ, *et al*. TheemerginglimbsandtwigsoftheEastAsianmtDNAtree. MolBiolEvol, 2002, 19(10):1737—1751.
22. MishmarD, Ruiz-PesiniE, GolikP, *et al*. NaturalselectionshapedregionalmtDNAvariationinhumans. ProcNatlAcadSciUSA, 2003, 100(1):171—176.
23. AchilliA, RengoC, MagriC, *et al*. ThemoleculardissectionofmtDNAhaplogroupHconfirmsthattheFranco-Cantabrianglacial l. AmJHumGenet, 2004, 75(5):910—918.
24. MacaulayV, HillC, AchilliA, *et al*. Single, rapid coastal settle mentof Asiarev ealed by analysisofcompletemitochondrialgenomes. Science, 2005, 308(5724):1034—1036.
25. IngmanM, KaessmannH, PaaboS, *et al*. Mitochondrialgenomevariationandtheoriginofmodernhumans. Nature, 2000, 408(6813): 708—713.
26. HerrnstadtC, ElsonJL, FahyE, *et al*. Reduced-median-networkanalysisofcompletemitochondrialDNAcoding-regionsequences forthemajorAfrican, Asian, andEuropeanhaplogroups. AmJHumGenet, 2002, 70(5):1152—1171.
27. HerrnstadtC, PrestonG, HowellN. Errors, phantomsandotherwise, inhumanmtDNAsequences. AmJHumGenet, 2003, 72(6):1585—1586.
28. Yao Yonggang, Kong Qingpeng, Sun Chang, *et al*. Can the truthfulness of base deletion/insertion in human mitochondrial DNA be inferred from the phylogenetic relationship of lineages?48(4):348—352.
29. YaoYG, MacauleyV, KivisildT, *et al*. Totrustornottotrustanidiosyncraticmitochondrialdataset. AmJHumGenet, 2003, 72(5): 1341—1346;authorreply1346—1349.
30. SilvaWA, Jr. , BonattoSL, HolandaAJ, *et al*. MitochondrialgenomediversityofNativeAmericanssup-

portsasingleearlyentryof founderpopulationsintoAmerica. AmJHumGenet, 2002, 71(1):187—192.
31. WallaceDC, BrownMD, LottMT. MitochondrialDNAvariationinhumanevolutionanddisease. Gene, 1999, 238(1):211—230.
32. WallaceDC. MitochondrialDNAsequencevariationinhumanevolutionanddisease. ProcNatlAcadSciUSA, 1994, 91(19):8739—8746.
33. TorroniA, CamposY, RengoC, *et al*. MitochondrialDNAhaplogroupsdonotplayaroleinthevariablepheno-typicpresentationof theA3243Gmutation. AmJHumGenet, 2003, 72(4):1005—1012.
34. YaoYG, KongQP, BandeltHJ, *et al*. Phylo geo graphic differen tiation of mitoch ondrial DNA in Han Chinese. Am JHum Genet, 2002, 70(3):635—651.
35. KongQP, YaoYG, SunC, *et al*. Phylogeog raphican alysiso fmito chon drial DNA haplogroup F2in China reveals T12338 Cinthe in itiation codon ofthe ND 5genenottobepathogenic. JHumGenet, 2004, 49(8):414—423.
36. TzenCY, ThajebP, WuTY, *et al*. MelaswithpointmutationsinvolvingtRNALeu(A3243G)andtRNAGlu(A14693g). MuscleNerve, 2003, 28(5):575—581.
37. Bandelt H J, Yao Y G, Kivisild T. Mitoch ondr ialgen esands chiz ophr enia. Schi zophr Res, 2005, 72(2-3): 267—269.