DETECTION AND ANALYSIS OF POLYMORPHIC NUMTs IN HUMAN POPULATIONS

PROJECTE DE RECERCA
MARIA GUARDIOLA RIPOLL. GRAU EN BIOLOGIA 2014

UAB Universitat Autònoma de Barcelona

INTRODUCTION

The process by which mtDNA sequences colonize the nucleus generates NUMTs –Nuclear Insertions of Mitochondrial Origin–. These insertions are widely distributed among all chromosomes and have been found in many eukaryotic organisms, including Human and other primates [1,2]. After they arrival to the nucleus, these insertions loss their function becoming pseudogenes [2]. The most ancient NUMTs are considerate to be "molecular fossils" [3].

It is believed that NUMTs get to the nucleus by NHED *(non-homologous end-joining)* major mechanism of double-strand break repair in mammal cells [4] avoiding chromosomic deletions [5, 6]. Once in the nucleus these insertions can suffer postinsertional changes like total or partial deletions or duplications. Duplicated NUMTs show homology in their flanking regions because of the lack of self-replication mechanism [7].

The colonization of NUMTs is an ongoing process providing important information of the recent human evolution and past demographic events [6].

INITIAL HYPOTHESIS AND OBJECTIVES

The main goal of this project is to describe Human NUMTs and obtain a database of Human Polymorphic NUMTs suitable to be used as genetic markers, a valuable tool that can provide new perspectives in the study of human population dynamics. Specific objectives are:

1. Update of NUMTs database published by Ramos et al., 2011a [8] with the new version of the human genome draft GRCh38 (hg20).
2. Classify NUMTs as human specific or non-human specific and distinguish between original insertions or duplication events.
3. Propose a methodology for NUMT dating.
4. Obtain a list of Human Polymorphic NUMTs to be used as genetic markers.
5. Validate NUMTs polymorphic nature in the laboratory.

Figure 1. How Human Polymorphic NUMTs must be:
[1]Insertion ocurred after *H.sapiens-P.troglodytes* divergence 4-6 Mya [6]
[2]Insertion ocurred after divergence of modern human 200,000-100,000 ya [9]

MATERIAL AND METHODS

TASK 1: Update NUMTs Database

1. Conversion of coordinates of 756 NUMTs described by Ramos et al., 2011a [8] to hg20 using Galaxy Suit [10,11,12].
2. Search of new NUMTs in hg20 version[1]. Reference sequences and BLAST tool available in NCBI.
3. Search of polymorphic NUMTs in 1000 Genomes Project Database[1] [13].

TASK 2: Classification[2] of NUMTs as human specific or non-human specific

1. Submission of NUMTs and corresponding flanking region to BLAST.
2. Inspection of generated alignments with BioEdit [15] and classify as Human Specific NUMT if it only appears in human lineage and as non-human specific NUMT if it appears in at least *Pan troglodytes* lineage.

Procedure to follow summarized in Figure 2.A, B, and C.

TASK 3: Classification[2] of NUMTs as original insertion or duplication

Classification of postinsertional processes is based on mtDNA region originating the NUMT.

1. Insertions that correspond totally or partially to the same mitochondrial region could represent a duplication and be consequence of a previous insertion event.
2. Follow procedure summarized in Figure 2.A and D.

TASK 4: Polymorphic NUMT selection

Analysis will be performed with NUMTs classified as Human specific whether are original insertions or duplications.

1. Follow procedure summarized in Figure 3, A.

TASK 5: Validation of Polymorphic NUMTs in laboratory

NUMT and flanking sequences analysis in laboratory by PCR and Agarose Gel Electrophoresis to see their polymorphic nature in human populations[5]. Procedure summarized Figure 3, B.

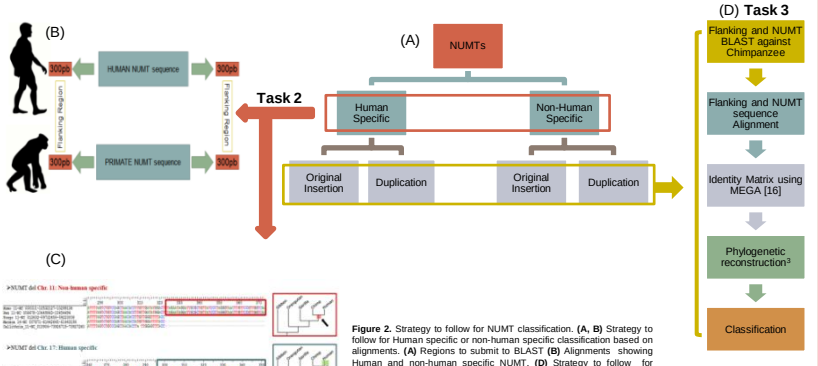

Figure 2. Strategy to follow for NUMT classification. **(A, B)** Strategy to follow for Human specific or non-human specific classification based on alignments. **(A)** Regions to submit to BLAST **(B)** Alignments showing Human and non-human specific NUMT. **(D)** Strategy to follow for classification as original insertions or duplications.

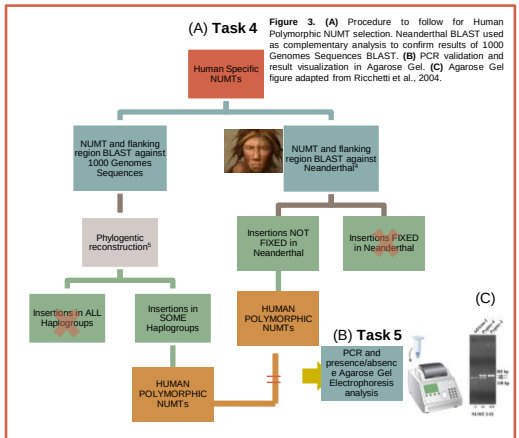

Figure 3. **(A)** Procedure to follow for Human Polymorphic NUMT selection. Neanderthal BLAST used as complementary analysis to confirm results of 1000 Genomes Sequences BLAST. **(B)** PCR validation and result visualization in Agarose Gel. **(C)** Agarose Gel figure adapted from Ricchetti et al., 2004.

[1]Ramos 20011a [8] methodology and Hazkani-Covo 2007 [7] criteria selection will be used for NUMTs searching.
[2]Procedures used for classification will be the same proposed by González M.M., 2011 [14]
[3]Mutation rate of human pseudogene will be used as Molecular Clock: 2.5 x 10^{-3} mutation per site per generation [17]
[4]Alignment using Neanderthal sequences available in UCSC [18] and chimpanzee as an outgroup
[5]Populations selected for PCR analysis will depend on data from task 4.

EXPECTED RESULTS

- It is expected to find the majority of insertions in human genome to be non-human specific in accordance with previous studies [6].
- It is expected to clarify the discrepancies about the origin of most insertions: whether the most are consequence of original insertion events [19], or due to further duplication events [2].
- It is expected the dating process to be useful to obtain, after laboratory validation, new genetic markers –Polymorphic NUMTs–.

TIMELINE AND BUDGET

The whole project is scheduled for a period of three years and a budget of 100,282.00 Euros (Table 1 and 2 respectively).

1.

TIMELINE	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.	
TASK 1													
NUMTs Database updating to hg31													
New NUMTs detection in hg20													2015
Polymorphic NUMTs in 1000 Genomes Project searching													
TASK 2													
Human or non-human specific NUMT classification													
Alignment inspection													
TASK 3													
NUMTs classification as original insertions or duplications													
Phylogenetic Reconstruction of duplication events													2016
TASK 4													
Phylogenetic Reconstruction of Human specific NUMTs													
BLAST with 1000 Genomes Project sequences													
BLAST with Neanderthal sequences													
TASK 5													
Polymorphic NUMTs validation													2016 2017

Table 1. Timeline and schedule.

2.

Disposable Material	45,000.00 €
Equipment	8,492.00 €
Bioinformatics	10,700.00 €
Research Personnel	25,000.00 €
Others	10,000.00 €
Total	**100,292.00 €**

Table 2. Project budget

REFERENCES

1. Bensasson et al. (2003). J. Mol. Evol. 57:343-354
2. Hazkani-Covo et al. (2003). J. Mol. Evol. 56:169-174
3. Hazkani-Covo et al. (2010). PLoS Genet. 6:e1000834
4. Woichnik and Moraes (2002). Genome Res. 885-893
5. Calabrese et al. (2012). BMC Bioinformatics. 13:S15
6. Ricchetti et al. (2004). PLoS Biol. 2:E273
7. Hazkani-Covo and Gaur (2007). Mol. Biol. Evol. 24:13-18
8. Ramos et al. (2011a). Mitochondrion. 11:946-953
9. Cooper and Kejrer-Sawatzki (2008). *Handbook of Human Molecular Evolution*. England: John Wiley & Sons Ltd.
10. Blankenberg et al. (2010). Curr. Protoc. Mol. Biol. 19:1-21
11. Giardine (2005). Genome Res. 15:1451-1455
12. Goecks et al. (2010). Genome Biol. 11:R86
13. Abecasis et al. (2012). Nature. 491: 56-65
14. González M.M. (2011). *Analysis of Nuclear Insertions of Mitochondrial Origin*. Màster en Biologia Humana. UAB.
15. Hall (1999). Nucleic Acids Symp. 41:13-18
16. Tamura et al. (2011). Mol. Biol. Evol. 28:2731-9
17. Nachman and Crowell (2000). Genetics 156:297-304
18. Kent et al. (2002). Genome Res. 12:996-1006
19. Lang et al. (2012). Hum. Genet. 131:757-771