論文の内容の要旨

論文題目

Excavating Ancient Mitochondrial DNAs from Mammalian Nuclear Genomes
（哺乳類核ゲノムにおけるミトコンドリアＤＮＡの包括的解析）

氏　　名　　辻　　淳　子

Mitochondria possess their own small genomes (around 16,000bp in mammals). NUMTs (Nuclear MiTochondrial sequences) are (partial) copies of the mitochondrial DNA (mtDNA) inserted into the nuclear genome via the repair of DNA double-strand breaks. Several computational studies have investigated NUMTs, however those studies have not used appropriate methodology for sensitive detection of NUMTs and precise delineation of their boundaries.

We developed a carefully considered protocol, *norg-suite*, to redefine NUMT datasets of four mammalian species (human, rhesus, mouse, and rat), and by analyzing the datasets, found new characteristics of NUMT integration sites. The issues we considered include appropriate alignment parameters, correct handling of circular mtDNA, masking of low complexity sequences, post-insertion duplication of NUMTs, long indels and validation of E-value thresholds. With simulated NUMT dataset, we confirmed that this method outperformed other methods used in previous researches. Since there is no tool which define NUMT region despite the fertility of NUMT studies to date, we packaged this protocol as a software tool to make it easy to excavate NUMT datasets of other species.

As the characteristics of NUMT insertion sites, we found four general features related to chromatin and genomic contexts in human, rhesus, mouse, and rat.

1　We resolved an outstanding controversy in the literature, by confirming that retrotransposons are highly enriched in NUMT flanks. NUMT insertion sites of all

four mammalian species show the significant over-representation of retrotransposons (binomial test, $p < 1 \times 10^{-4}$).

2 We discovered that NUMT insertion sites show a marked tendency to have high DNA curvature. In all organisms, NUMT flanks in the first 20bp showed significant high predicted DNA curvature.

3 We found that each species show a significant enrichment of A+T oligomers in the first 10bp of NUMT flanks: TATATA ($p \sim 4.2 \times 10^{-4}$) in human, ATTATT ($p \sim 7.9 \times 10^{-8}$) in rhesus, AAACTT ($p \sim 1.1 \times 10^{-4}$) in mouse, AATTTA ($p \sim 4.4 \times 10^{-4}$) in rat. Interestingly, the oligonucleotide recognized by L1-endonucleases (TTTTAA) was also significantly enriched ($p \sim 1.4 \times 10^{-3}$).

4 We quantified the degree to which NUMT insertion sites correspond to open chromatin regions identified by the DNaseI-seq and FAIRE-seq experimental methods. By cross-checking with the open chromatin data, we found that NUMT insertions correlate with open chromatin regions in most measured cell types (binomial test, $p \ll 0.05$).

More interestingly, we show that the correlation between NUMTs and open chromatin regions drops sharply when examining NUMTs older than the split between human and chimpanzee. This suggests that open chromatin regions where NUMTs insert (primarily intergenic regions) has shifted dramatically in the human line during relatively short evolutionary time scales. In addition to this, we observed that species-specific NUMT insertion sites are enriched in species-specific open chromatin regions. This highlights the possibility of the utility of NUMTs as open chromatin markers.