

認識粒線體假基因

黃家富、劉富光

水產試驗所淡水繁養殖研究中心

前言

粒線體是直接提供細胞生命活動能量的地方；其含有染色體 DNA 以外之遺傳物質，稱之為粒線體 DNA (mtDNA)。每一細胞體中含有數個 mtDNA，其具：(1)分子量小、約 15–18 Kb，分子結構為一個閉合環狀雙鏈分子，易於分析；(2)母系基因遺傳 (Maternal inheritance) 的特性；(3)不具修補系統 (Repair mechanism)，複製過程易發生變異，演變速率較染色體 DNA 為快；(4)基因組成相當穩定，含有 13 個編碼蛋白質基因 (Protein coding gene)、2 個核糖體 RNA 基因、22 個轉移 RNA (tRNA) 及一無結構基因 (稱為控制區，Control region 或 D-環，D-loop)，且不含有介入子 (Intron)，非密碼區很少等特性；因此，結合 mtDNA 標記和以聚合酶鏈鎖反應 (PCR) 為基礎的定序技術，廣泛運用於系統發育學、族群遺傳、族群結構及類緣關係等研究。

Numt

然而在真核生物中的染色體基因組中，存在一些由粒線體基因組轉移進入染色體基因組中的 DNA 片段或是相似的細胞核 DNA

片段，這些被認為是分子化石的片段稱之為“粒線體 DNA 核內插入序列 (nuclear insertion of mitochondrial DNA, Numt)” (Lopez et al, 1994)，或者稱為“粒線體假基因 (nuclear mitochondrial-like sequence)” (Wang, 2004)。早在 1967 年，Dubuy 和 Riley 首次發現 Numt，可是遺憾的是，在 90 年代以前，這問題一直沒受到學者的關注，直到 1994 年，Lopez 等用 Southern 雜交的方法在家貓染色體基因中發現，mtDNA 從調控區 (D-loop) 到 CO II 基因完整的大片段轉移，通過序列比對發現的粒線體基因同源序列；其後許多學者在多種動植物的染色體基因組內都發現存在 Numt 現象。mtDNA 與染色體基因組發生非同源重組是 Numt 產生與積累的重要原因，而染色體基因組倍增是 Numt 數目增多的另外一個原因，因此 Numt 在不同物種中積累的數目差異很大，在人類、水稻及擬南芥中存在大量的 Numt，但在秀麗廣杆線蟲、原雞和蜜蜂中卻存在很少的 Numt 或者甚至沒有，造成這種數目差異的原因，目前並不清楚。

在 mtDNA 研究中的應用已有一系列綜述性文獻，雖其涉及到幾個相關的問題，包括父系遺傳、異質性、進化異速性和非中性等，但是在 mtDNA 分析中，為粒線體基因

的旁系同源物-Numt 的排除成為不可避免的首要解決的問題。目前文獻顯示 Numt 具有幾個特點：(1)染色體基因組中的 Numt 序列與其 mtDNA 同源序列相比，進化模式完全不同，由於不受選擇壓力等影響，Numt 序列常發生插入或缺失，造成移框突變，甚至發生倒位或易位現象；(2)Numt 與粒線體基因比較，其序列的進化速度相對較慢，所以常被認為 mtDNA 序列的分子化石，因此當使用物種間的通用引子或者基於保守區域設計的引子時，Numt 更容易被 PCR 擴增，在對 PCR 產物進行電泳分析時，往往有非特異性條帶出現，增加背景雜訊，對測序非常不利，不僅增加研究工作的難度，必然會混淆後續的資料分析，甚至得出錯誤的物種間進化關係和生物系統地理學結論；(3)有些 Numt 序列可作為獨立元件在染色體基因中複製、移動。

隨著基因組序列的測定，對不同物種 Numt 系統分析的工作已經逐步開展。人類基因組的分析顯示，最長的 Numt 片段達到 14,654 bp，片段長度大於 800 bp 的數目為 110 個。由於以往未對人類 Numt 進行系統分析，Numt 富集現象造成 PCR 過程中的 Numt 混入，因此得出錯誤的疾病研究結果，甚至是人類 Numt 被錯誤認為是恐龍的 DNA 序列。Zhang et al. (2006) 以中國東南沿海的擬穴青蟻 (*S. paramamosain*) 為研究對象，以粒線體 COI 基因的通用引子和特異性引子進行擴增，分別得到 34 個假基因和 5 個粒線體 COI 基因序列，在 34 個假基因中共定義了 29 種單倍型，根據序列的相似度，這些假基因可以分為 2 類，每類假基因

都有各自保守的核苷酸序列，第 I 類假基因存在 2 處插入序列和 1 處 8 bp 的缺失序列，這些位點導致了整個閱讀框的移位；在第 II 類假基因和 5 個粒線體 COI 序列中只有基替換，未發現插入和缺失序列。研究分析表明，這兩類假基因分別代表了 2 次核整合事件；故利用 mtDNA 進行青蟹遺傳結構和親緣演化研究時應警惕假基因的干擾。

Numts 避免與檢查方法

要如何避免和檢查 Numts 的主要方法有以下幾種：

一、擴增組織的選擇

採用 mtDNA/核 DNA 比值高的組織或擴增前純化粒線體，可以避免 Numt。Numt 在細胞核中有相當高的拷貝數，相同的 PCR 引子則更易與 Numt 結合，擴增 Numt 序列片段。當使用與核 DNA 相比更富含 mtDNA 的組織—如肝臟、心臟或生殖腺，在製備 mtDNA 前先純化粒線體，則可得到良好的模板，避免擴增到 Numt 片段。

二、設計特異性引子

在序列分析比較研究中，引子設計至關重要；由於 Numt 與 mtDNA 相比進化速度更慢，同祖先序列相比更相近，並在相關分類單元間更相似，因此，根據特定物種序列設計引子，則不易擴增出 Numt 序列；若使用非特異 PCR 引子 (以物種分類單元設計的引子和通用引子) 時，由於引子與核拷貝有更好的配置，因而 Numt 易於優先擴增，這已在嚙齒目動物、靈長類動物等的研究中，使用粒線體 *Cyt b*、12s rRNA 和 D-loop 序列

作為診斷標記得到證實。理想的定序方案是根據定序片段的末端序列來設計特異性引子，但首先應做的工作是純化 mtDNA。

三、反轉錄 PCR

Blanchard et al. (1996) 認為 Numt 能以反轉錄 PCR (RT-PCR) 技術來避免，可惜仍偶爾有 Numt 被轉錄。當知道 Numt 和 mtDNA 序列，且 mtDNA 是單系發生的，就可設計特異性的引子。

四、以長鏈 PCR 產物來排除短而模糊的 Numt

曾有學者認為擴增整個 mtDNA 基因或大部分 mtDNA，是避免 Numt 一有效的方法；但隨著對 Numt 序列之研究，此一說法已引起爭議，認為此理論缺乏證據。

五、Numt 和 mtDNA 被同時擴增後的分離方法

當 Numt 和 mtDNA 被同時擴增出來時，有一些技術可以把它們分離：(1)PCR 產物先被連接到載體上選殖 (clone)，然後對不同的選殖進行定序；(2)選擇在兩條序列變異最大的地方，設計內部引子去區分兩條不同的序列，最理想的情況是，新引子的 3' 端鹼基最好選擇 mtDNA 和 Numt 序列發生顛換位置的鹼基或至少是不同的位置，這種方法對於研究種內變異比選殖定序更有效；一旦得到兩條清楚的 DNA 序列，就可以根據討論的標準區分哪個是 Numts，哪個是 mtDNA。

六、使用純化 mtDNA 作探針進行南方雜交 (Southern hybridization)

一個更直接但技術上較困難的 Numt 檢測方法是使用純化 mtDNA 作探針進行南方

雜交。首先從血樣或其他組織抽提的 DNA 用限制性內切酶消化，通過瓊脂膠電泳使片段分離，然後轉移到硝酸纖維素膜或尼龍膜上。在缺乏 Numts 的樣品中，純化的粒線體探針雜交後僅僅檢測到 mtDNA 陽性帶。如存在 Numts，則額外的染色帶將被檢查到。即使 Numts 具有同 mtDNA 一樣的限制性內切酶位元點，但在末端區域應生成特異長度的條帶，這些限制性條帶可能就是 Numts 串聯重複區域。

結語

粒線體假基因的存在給 mtDNA 標記的系統發育、族群遺傳研究帶來麻煩，但同時由於其基因保存著祖先功能基因的殘餘拷貝，可能成為研究生物進化和基因組動態變化，分析基因複製與突變等事件的年代及頻率，揭示基因替換、插入與缺失等事件的機制提供重要線索，有助於更進一步理解進化本質，所以假基因並非基因組中的垃圾序列，而是具有潛在性的基因序列。