

致謝

在這兩年短短時光裡，從一個從來沒有接觸過實驗室的我，開始學習閱讀文獻、PCR、抽 DNA 到寫程式解決問題等，這一路上我很慶幸自己能遇到一位令人敬愛的好老師邱偉欣。他對我們的管教從來不是時時刻刻的盯緊進度，而是放手讓自己去安排和規劃實驗，讓我們學習獨立和自立自強。每當有疑問產生的時候，老師也不會直接將答案說破，而是循循善誘的引導我們思考的邏輯和方向。此外，老師值得我們尊敬的地方除了在課業外，還有他對待流浪動物那份心意，不論是在照顧或醫療上給了我許多的幫助和建議，無盡的感動和謝意盡在不言中。甚至對於我在研究及生活上的任性和沮喪，老師也會不時的展現出他的寬容並且教導及指正我的錯誤。邱老師或許是我在學校生涯中的最後一位老師了，然而他的寬大和對教學的熱忱卻深深的感動了我。這兩年的點點滴滴和滿滿的謝意，再多的話到了這裡也只是簡單的三個字，「偉欣老師，謝謝您！」。沒有您的包容與教導就不會有現在的我。

此外，也謝謝陳荷明和謝明麗老師對我在研究和論文上的建議及指導，以及感謝所有實驗室的夥伴奕憲、敘竹、又新、明源和清平一路上的陪伴和相助，尤其是淑閔，感謝她與我一同歡笑與哭泣，也分擔了所有的挫折與喜悅。感謝東海大學生科的所有同學們雅玄、閔瑄、

柏凱、祐萱和依庭，在這兩年的時光中陪我一起成長、茁壯和揮灑青春。最後一定要感謝我的家人們，因為他們的支持才能使我無後顧之憂的完成我的碩士班研究。而如今要畢業了，對於大家還是只有那一句說不膩的「謝謝」。

目錄

	頁次
目錄	I
圖目錄	III
表目錄	V
附錄	VI
中文摘要	VIII
英文摘要	X
前言	1
材料與方法	7
壹、硬體	7
貳、軟體	7
參、原核生物序列特徵的資料庫建立	8
肆、真核生物序列特徵的資料庫建立	10
伍、原核生物測試序列	11
陸、真核生物測試序列	13
柒、資料庫以外的測試序列	13
捌、對序列特徵資料庫的分類結果進行比對	15

	頁次
結果	16
壹、RAIphy 軟體測試	16
貳、 <i>Prevotella</i> 屬序列的分類結果	17
參、測試序列 <i>Lactococcus garvieae</i> Lg2 的分類結果	23
肆、原核生物序列特徵資料庫測試	23
伍、真核生物序列特徵資料庫測試	26
陸、混合序列測試	27
柒、非資料庫序列測試	30
討論	32
參考文獻	34
圖	38
表	94
附錄(存於光碟中)	

圖目錄

	頁次
圖一、RAIphy 測試-各物種在屬層級的分類準確度	38
圖二、RAIphy 測試-屬層級分類的平均準確度	39
圖三、 <i>Prevotella</i> 屬的三個物種的在各層級的分類結果	40
圖四、 <i>Prevotella denticola</i> 在門層級的分類結果	41
圖五、 <i>Prevotella denticola</i> 在綱層級的分類結果	42
圖六、 <i>Prevotella denticola</i> 在目層級的分類結果	43
圖七、 <i>Prevotella denticola</i> 在科層級的分類結果	45
圖八、 <i>Prevotella denticola</i> 在屬層級的分類結果	48
圖九、 <i>Prevotella ruminicola</i> 在門層級的分類結果	52
圖十、 <i>Prevotella ruminicola</i> 在綱層級的分類結果	53
圖十一、 <i>Prevotella ruminicola</i> 在目層級的分類結果	54
圖十二、 <i>Prevotella ruminicola</i> 在科層級的分類結果	55
圖十三、 <i>Prevotella ruminicola</i> 在屬層級的分類結果	58
圖十四、 <i>Prevotella melaninogenica</i> 在門層級的分類結果	62
圖十五、 <i>Prevotella melaninogenica</i> 在綱層級的分類結果	63
圖十六、 <i>Prevotella melaninogenica</i> 在目層級的分類結果	64
圖十七、 <i>Prevotella melaninogenica</i> 在科層級的分類結果	66

	頁次
圖十八、 <i>Prevotella melaninogenica</i> 在屬層級的分類結果	68
圖十九、 <i>Lactococcus garvieae</i> Lg2 在門層級的分類結果	71
圖二十、 <i>Lactococcus garvieae</i> Lg2 在綱層級的分類結果	72
圖二十一、 <i>Lactococcus garvieae</i> Lg2 在目層級的分類結果	73
圖二十二、 <i>Lactococcus garvieae</i> Lg2 在科層級的分類結果	75
圖二十三、 <i>Lactococcus garvieae</i> Lg2 在屬層級的分類結果	78
圖二十四、無質體的原核生物序列特徵資料庫測試	83
圖二十五、有質體的原核生物序列特徵資料庫測	84
圖二十六、原核生物資料庫的平均分類準確度	85
圖二十七、三十種原核生物序列測試	86
圖二十八、三十種真核生物序列特徵資料庫測試	87
圖二十九、混合序列測試真核生物序列特徵資料庫	88
圖三十、比較混合與單獨使用真核生物序列的分類結果	89
圖三十一、序列測試未加質體的原核生物序列特徵資料庫	90
圖三十二、比較混合與單獨使用原核生物序列的分類結果	91
圖三十三、三十個種層級不存於資料庫的序列類測試	92
圖三十四、三十個種層級不存於資料庫的序列類測試	93

表目錄

	頁次
表一、原核生物序列特徵資料庫所包含的分類層級	94
表二、真核生物序列特徵資料庫所包含的分類層級	101
表三、測試序列的物種名稱	102
表四、 <i>Prevotella denticola</i> 在各分類階層的分佈情形	103
表五、 <i>Prevotella ruminicola</i> 在各分類階層的分類情形	105
表六、 <i>Prevotella melaninogenica</i> 在各分類階層的分類情形	107
表七、 <i>Lactococcus garvieae</i> Lg2 在各分類階層的分類情形	109
表八、不同操作方法是否影響無質體序列資料庫的分類準確度	112
表九、不同操作方法是否影響真核資料庫的分類準確度	113
表十、不同操作方法是否影響原核資料庫分類混合序列的結果	114
表十一、比較混合與真核序列的分類結果	115
表十二、不同操作方法是否影響原核資料庫分類混合序列的結果	116
表十三、比較混合與原核序列的分類結果	117
表十四、不同操作方法是否影響不包含於資料庫的種序列分析	118
表十五、不同操作方法是否影響不包含於資料庫的屬序列分析	119

附錄

*附錄內容附於光碟中

附錄一、complete-choose-separate.py

附錄二、complete-genome1129.py

附錄三、gen-str1215-2.py

附錄四、fasta-1.py

附錄五、separate-pla1216.py

附錄六、complete-pla-pla-link.py

附錄七、test-pla1213.py

附錄八、test-pla,1213.py

附錄九、seqlink.py

附錄十、complete-choose0117.py

附錄十一、complete-choose0201.py

附錄十二、Delmito0131.py

附錄十三、plasmid-separate.py

附錄十四、DelContigClone0206.py

附錄十五、taxa-search0309.py

附錄十六、rancircle.py

附錄十七、ran.py

附錄十八、fungi-CandiadaA.py

附錄十九、split-gi-id.py

附錄二十、len-check.py

附錄二十一、m-taxasearch67-128.py

附錄二十二、m-taxasearch128-188.py

附錄二十三、m-taxasearch188-248.py

附錄二十四、m-taxasearch248-254.py

附錄二十五、delrepeattaxa.py

附錄二十六、taxa-search.py

附錄二十七、taxa-search-1.py

附錄二十八、taxa-search-2.py

附錄二十九、contiglenchoos-sp.py

附錄三十、contiglenchoos-ge.py

附錄三十一、alignGenus.py

附錄三十二、taxa0423pro.py

附錄三十三、taxa0407euk.py

中文摘要

多源性基因體學是研究某一特定環境中所有微生物的基因體的一門學科。它的研究方法特點在於，不經微生物培養的方式而直接萃取環境微生物的基因體，再利用高通量定序法定序，然後利用生物資訊學的手段進行分析以瞭解來自環境微生物的基因體所承載的訊息。而在這些分析當中，了解環境微生物的多樣性是至關重要的。序列特徵廣泛的被用於分類原核生物序列，因為，特定樣式的寡核苷酸序列在不同物種的基因體中出現的頻率不同，並且是均勻分布在基因體中的。先前的文獻指出，真核與原核生物的基因體有類似的序列特徵，但目前大多數多源性基因體學，均偏重環境中原核生物的多樣性分析而缺乏針對環境中真核生物多樣性的探討。RAIphy 是被設計用來分類多源性基因體資料中原核生物序列片段的一種軟體，且相較於其他軟體分類準確度較好；該軟體是利用已知的基因體序列特徵，建立出一個分類模型進行分類。有鑑於真核生物與原核生物皆俱有相似的基因體序列特徵，因此在我的研究中，我嘗試利用 RAIphy 這個軟體來分類真核生物的 DNA 片段。從目前的結果中發現，利用序列特徵確實可以將真核生物的序列片段分類到正確的階層 (taxa)，但是混合真核生物與原核生物的序列來進行分類時，分類的準確程度卻明顯的下降。這些結果顯示，原核生物與真核生物基因體上的序列特徵可能是十分

相近的；而這也可能是由於 RAlphy 所採用的演算法無法清楚分類真核生物和原核生物的序列特徵，有鑒於此，開發新的演算法是必要的。

Abstract

Metagenomics aims to identify microbes or genes in a specific environment by sequencing DNA fragments directly followed by bioinformatic analyses. Binning, which aims to classify DNA fragments to corresponding taxonomic ranks, is a critical step to understanding species diversity in an environment. Composition-based binning methods, which apply genomic signatures to classify DNA fragments, are the state-of-the-art ways to explore species diversity in metagenomes, since genomic signatures are supposed to be unique to species and distribute homogeneously in a genome. RAlphy is software designed to classify prokaryotic sequences in metagenomic data by using k-mer genomic signature. Previous studies showed that eukaryotic genomes seem to have similar genomic signatures to that in prokaryotic genomes. Nevertheless, recent metagenomic studies made more efforts to study environmental prokaryotes than eukaryotes. In this study, we tried to use RAlphy to classify both prokaryotic and eukaryotic DNA fragments based on their genomic signatures. The results showed that both prokaryotic and eukaryotic DNA fragments could be classified to their own taxa with high accuracy rates, while mixed sequences of both prokaryotes and eukaryotes caused a decrease of accuracy rate. These results suggested that eukaryotes should have very similar genomic signatures to prokaryotes and a novel algorithm to classify eukaryotic and prokaryotic DNA fragments is expected.

前言

多源性基因體學(Metagenomics)是研究某一特定環境中，該環境內所有微生物基因體的一門學科(Huson et al.,2007)。其取樣的環境可以是海洋、土壤或是動物的腸胃道等等，只要有微生物存在的地方，都可以當作取樣的地點。而這些從環境中所取得的樣本會經由初步過濾後，將其 DNA 給萃取出來，利用限制內切酶或超音波震盪的方式將 DNA 切碎或震碎成較短的片段，這些短片段的 DNA 則會進一步定序，而定序完成的微生物基因體片段，就稱為 reads。在進行接下來一連串的生物資訊分析之前，可以先將 reads 組裝成較大片段的 contigs，或是直接分析短片段的 reads。而組裝後的序列在分析上雖然可得到較多較為完整的的資訊，但卻有幾項缺點，一是在組裝序列的過程中，不可避免的會出現錯誤，一旦組裝發生了錯誤也會連帶影響之後所有的分析結果。二是在有些多源性基因體學的資料中因為生物多樣性太高，根本無法將這些定序出來的 reads 組裝成較大片段的 contigs (Tringe and Rubin, 2005)。第三若是環境內的生物其各個族群的個體數目太過於平均無明顯地落差的話，也會很難將 reads 組成較大片段的 contigs。因此若是可以直接對短片段的 reads 進行分析就可以避免上述的錯誤產生。

生物資訊學在多源性基因體學上的分析，主要包含了解該環境內

微生物的多樣性及對其基因功能的研究與註解等等。前者可以幫助了解該環境內微生物的組成；後者則可以了解該環境內的微生物主要帶有那些特定的基因，以及這些特定的基因是否與這些微生物適應該環境有所關係。

將 reads 或 contigs 分類至其所屬的分類階層的這個過程，稱為 binning(Wooley et al., 2010)。Binning 據其分類的方式可分為兩種，第一種是利用同源序列的搜尋將序列加以分類和註解 (similarity-based binning)，常用的工具如 blastx (Altschul et al., 1990)。雖然這種類型的 binning 方式可以得到較高的準確度，但卻存在幾個缺點，一若是資料庫中缺乏欲比對序列的同源序列時，便無法利用此法進行比對和分類 (Dinsdale, 2008)；二是當欲分類的序列長度越短，其在與資料庫中的序列比對時，就越容易比對出多條親緣關係相似度很高的序列，進而造成準確度下滑(Krause, 2008)。另一種分類序列的方式，則是利用序列特徵作為其分類的依據。序列特徵最早是在 1960 年代由一位生化學家所發現，他發現當利用兩個單核苷酸出現在一段序列的頻率除以這兩個核苷酸一起在該序列出現的頻率時($f(XY)/f(X)f(Y)$, $X,Y=\{A,G,C,T(U)\}$)，可以得到一個數值 ρ ，而同一段序列會出現十六個不同 ρ 值，最大的 ρ 值對於不同的物種來說有其特異性，可以做為物種與物種間辨識的依據(Josse et al., 1961)。受限於當時的定序技術並

不發達，序列特徵的發現在 1970 年代並無得到蓬勃的發展。在 1980 年代的時候 Karlin 和 Burge 兩位科學家證明了兩件事，一是序列特徵是均勻的分布於該物種的基因體序列上，二是序列特徵是存在於所有生物之中(Karlin and Burge, 1995)。然而序列特徵發展至今，並不單只是看兩個核苷酸出現的頻率，甚至還包含了四個到六個核苷酸出現的頻率等等，只要其符合兩個條件即可，一是該序列特徵在其基因體序列中是均勻分布的，二是該序列特徵在物種間是有特異性的。因此，若是利用序列特徵的方式去分類序列至其分類階層，相較於序列相似性的比對來說，較不受限於資料庫中序列種類之多寡所影響 (Nalbantoglu et al.,2011)。

目前而言，要使用序列特徵來進行多源性基因體片段的 binning，可藉由機器學習 (machine learning) 的方式來達成。機器學習的演算法分為三種類型，unsupervised, supervised 以及 semisupervised。

Unsupervised 的原理是利用機器直接去辨識欲分類的序列其序列特徵後再進行分群，其操作方式有下列兩種，一是先在欲分類的序列中混合一些已知的序列，再藉由分群結果中的已知序列判斷原本欲分類的序列其分類階層；二則是直接讀取欲分類序列的序列特徵後進行分群，再從分群的結果中找出是否有標記序列存在(如 16S rDNA)進行判斷。當欲分類的序列中有從未曾被發現過的序列存在時，便可透過

unsupervised 的分群方式進行分類(McHardy et al.,2007)。但因為欲分類的序列在定序和組裝序列的過程中都有可能產生錯誤，因此若是欲分類的序列在被分類時沒有較為完整的已知序列的序列特徵作為其辨識的依據，那麼這些序列在分類結果的準確度上便不會太精確(McHardy et al.,2007)。Supervised 相較於 unsupervised 在分類序列上則較為精確，其原理是先利用機器去學習資料庫中已知序列的序列特徵以建立一個分類序列的模型，再藉由這個模型去辨識和分類序列，而此法因為在分類上有較為完整的已知序列的序列特徵作為其辨識的依據，因此在分類結果的準確度上較 unsupervised 精確(McHardy et al.,2007)。Semi-supervised 則融合了 supervised 與 unsupervised 的原理，先讓機器學習辨識資料庫中已知序列的序列特徵後建立模型，當欲分類的序列被放入分類模型後，機器便會將這些序列的序列特徵進行記憶並修飾原先的分類模型，最後再以被修飾過後的分類模型進行序列的分類，如此一來便可以將原先資料庫中未知序列的序列特徵也一併加入，使得其分類的準確度上較前面兩個方法佳(Nalbantoglu et al.,2011)。

目前多源性基因體學的研究集中於研究原核生物，對於環境中的真核生物則探討較少，但已知無論是在一般的海水中或是在極端的溫泉環境都可發現真核生物的存在(Aguilera et al., 2010、Massana, and

Pedr'os-Alio, 2008、Richards and Bass, 2005), 我們知道欲了解一特定環境內的所有微生物的組成及這些微生物的基因功能與環境間的關係的話, 是不可以忽略真核生物的存在的。早期的時候, 序列特徵主要是在原核生物上先開始研究的, 隨著定序技術的演進, 發現真核生物也有類似於原核生物的序列特徵, 只是原核生物的基因體序列上只會出現一種序列特徵, 但對於部分真核生物來說則可能會出現多於一種以上的序列特徵(Chor et al., 2009)。因此我們想嘗試是否可以利用序列特徵的方式去分類多源性基因體學資料中真核生物的序列片段。但在這之前, 我們需要選擇一個合適的軟體去分類真核生物的序列片段。

在眾多的利用序列特徵去分類多源性基因體學序列片段的軟體中, RA1phy 就目前來說是最為合適的軟體。以分類的序列長度來說, RA1phy 對於分類 400bp 長短的序列片段, 至少還有大於 60% 的準確度, 並且在操作上可以選擇使用 supervised 或 semi-supervised 兩種方式。對於序列特徵資料庫的建立也可依據使用者的需求自行建立與操作, 因此我們便可以利用 RA1phy 建構真核生物序列特徵資料庫進行一連串的分析 (Nalbantoglu et al., 2011)。首先我們利用 RA1phy 軟體分別建構了原核與真核生物的序列特徵資料庫, 並模擬多源性基因體資料中的 reads, 其片段大小為 500bp, 對建構好的序列特徵資料庫進行

測試。從目前結果得知真核生物與如原核生物的序列一樣，皆可以利用序列特徵的方法將真核生物的序列片段分類至其所屬的分類階層。此外，也證明了利用 semi-supervised 的方法去分類多源性基因體學的資料會是個合適的方法。

材料與方法

壹、 硬體

所使用的伺服器型號為 HP ProLiant ML110 G3，有四核心處理器、記憶體容量 8G，為 linux 操作系統。

貳、 軟體

RAIphy (Nalbantoglu et al.,2011) 是被設計用來分類多源基因體學資料中基因體片段的一種軟體，利用序列特徵 k-mer 作為其辨識的依據。RAIphy 可透過 supervised 和 semi-supervised 兩種機器學習的方式達成分類序列的目的，前者是讓機器去記憶已知序列的序列特徵，之後機器再依據記憶將未知序列做分類。後者則是除了記憶已知序列的序列特徵外，也會記憶未知序列的特徵，將原先的分類模型做一些修飾後，再將未知序列的序列予以分類，如此一來可得到較高的分類準確率。

在實驗處理的部分，皆以程式語言 Python(Bassi, 2007)做處理。Python 是一種直譯式的電腦語言，相較於其他軟體它較為簡易、好操作。

參、 原核生物序列特徵的資料庫建立

1. 原核生物的序列來源：建構資料庫的原核生物序列皆從 US National Center for Biotechnology Information (NCBI)的序列資料庫 (Pruitt et al.,2009) 中下載 (ftp://ftp.ncbi.nih.gov/refseq/release/)。而下載的原核生物序列所呈現的是 FASTA 格式，每七十個核苷酸為一行，但因顧及在接下來的步驟中序列操作上的方便，這些 FASTA 格式的核苷酸序列會利用程式的操作，將單行序列連接成一行。我們期望所建構的序列特徵資料庫其序列特徵是源自於完整的序列上，所以從下載的原核生物序列中挑選出有完整被定序的基因體序列 4036 條，這 4036 條完整的原核生物序列中，包含了質體與染色體序列。為了建構有加入質體序列以及沒有加入質體序列這兩種形式的原核生物序列特徵資料庫，須進一步地將質體序列與染色體序列分開，分開後一共得到 1456 條染色體序列及 2472 條質體序列，上述步驟皆由程式 complete-choose-separate.py(附錄一)所完成。
2. 建構未加入質體序列的原核生物序列特徵資料庫：因為有些物種可能有多條染色體序列的情形存在，因此先利用程式 complete-genome1129.py (附錄二) 將前步驟所挑出的 1456

條染色體序列分成兩類，一類是該物種的染色體序列多於一條以上，另一類則是該物種的染色體序列只有一條，前者會繼續利用相同的程式 `complete-genome1129.py` 將同一物種的不同條染色體序列相連一起，而後者因為有些序列有重複的資料(即相同物種的同一條染色體源於不同出處，而其序列上只有些微不同者，我們認定其為重複資料)，針對這個情形會利用程式 `gen-str1215-2.py`(附錄三) 選用出重複序列中的一條當作代表。最後會利用 linux 合併檔案的指令(`cat`)，將上述連接好的染色體序列檔案以及刪除重複序列的染色體檔案兩者合併後，再利用程式 `fasta-1.py` (附錄四) 將檔案中的序列轉換成每七十個核苷酸為一行的 FASTA 格式，交給 RA1phy 軟體讀取以建立未加入質體序列的序列特徵資料庫。

3. 建構有質體序列的原核生物序列特徵資料庫：在有加入質體序列的原核生物序列資料庫的建構上，須將質體序列與宿主細胞的染色體序列連接在一起，為了達成此目的，第一步先將質體序列分成兩類，一類是同物種中有多個質體序列者；另一類則是同一物種中只有一個質體序列者，此步驟參考程式 `separate-pla1216.py`(附錄五)。前者需先將相同宿主的多條質體序列利用程式 `complete-pla-pla1216.py`(附錄六)進行連接，後者

則不需。接下來再將前步驟的兩類質體序列分別連接於宿主細胞的染色體利用程式 test-pla1213.py、test-pla,1213.py (附錄七、附錄八)，而連接完的序列檔案會再利用程式 fasta-1.py (附錄四) 將檔案中的序列轉換成每七十個核苷酸為一系列的 FASTA 格式，再置入 RAlphy 軟體中被讀取，並建構加入質體序列的原核生物序列特徵資料庫。不論是有加入質體序列或是未加入質體序列的原核生物序列特徵資料庫，皆包含了三十一個門、五十個綱、一百零九目、兩百二十四科、五百四十四屬和一千零二十五個種(表格一)。

肆、 真核生物序列特徵的資料庫建立

1. 真核生物序列來源：建構資料庫的真核生物序列皆從 US National Center for Biotechnology Information (NCBI) 的序列 (Pruitt et al., 2009) 資料庫中下載 (<ftp://ftp.ncbi.nih.gov/refseq/release/>)，下載的真核生物序列包含了植物、無脊椎動物、原生動物、真菌和脊椎動物。
2. 真核生物染色體序列篩選：首先先利用程式 seqlink.py (附錄九) 將 FASTA 格式的真核生物序列連接成單一序列後，在接下來便會進行一連串序列篩選。第一步先利用程式

complete-choose0117.py (附錄十)挑選出植物、無脊椎動物和原生動物序列資料中有完整被定序出的基因體序列。然而因為真菌、脊椎動物等的序列資料較為龐大，因此使用程式 complete-choose0201.py (附錄十一) 將其序列資料中有完整被定序的基因體挑選出來。而這些有被完整定序的真核生物基因體會進一步的利用程式 Delmito0131.py (附錄十二) 將其內的粒腺體序列給去除，並用程式 plasmid-separate.py (附錄十三) 去除質體序列，最後再用程式 DelContigClone0206.py (附錄十四) 將 contig 片段序列去除。將最後篩選出的物種序列其名稱紀錄後，利用程式 taxa-search0309.py (附錄十五) 去 NCBI 網站將每個物種的分類階層紀錄下來，最後再將序列檔案置入 RA1phy 中讀取以建構真核生物列特徵資料庫。而真核生物序列特徵資料庫一共包含了九個門、十四個綱、十五個目、二十一個科、三十三個屬和三十九個種(表格二)。

伍、 原核生物測試序列

1. 為了知道 RA1phy 所提供的原核生物序列特徵資料庫，是否只能分類它所提供的原核生物測試序列，或是其他序列亦可？因此我們選擇了二十個不同 strain 的人體腸道細菌染色體序列

(Turnbaugh et al.,2009) (表格三)。這些序列皆從 NCBI 資料庫中下載，並利用程式 seqlink.py(附錄九)將 FASTA 格式的各個原核生物序列分別連接成一條序列的形式後，再利用程式 rancircle.py(附錄十六)模擬原核生物的環狀染色體的狀態，使其隨機切割 500bp 的長度(切 500bp 的長度是為了要模擬多源性基因體學定序資料的 DNA 長度片段)，切至每種序列各有五千條為止，利用 linux 的指令(cat)將這些五千條的序列檔案合併成單一檔案，因此共得了 10 萬條的原核生物測試序列。

2. 欲知道原核生物序列特徵資料庫在分類原核生物序列時除了屬之外，在不同的分類階層中其分類的準確度為何？因此利用前述的原核生物序列再進行一次測試，但因前面所提及的二十個不同 strain 的人體腸道細菌染色體序列中，包含了同個物種但不同 strain 的序列，而為了使得測試的序列能涵蓋更多樣的物種，並測試在不同的分類階層中分類的準確度。因此將前面不同 strain 但同物種的序列擇一留下後，再增加十四種(表格三)人體腸道細菌染色體序列，最後共得三十種(表格三)序列 (Turnbaugh et al.,2009)。而這十四種新加入的人體腸道細菌染色體序列，皆會從 NCBI 資料庫中下載，利用程式 seqlink.py(附錄九)將 FASTA 格式的原核生物序列連接成一條序列的形式後

(同前)，再利用程式 `rancircle.py`(附錄十六)模擬原核生物的環狀染色體的狀態，隨機切割 500bp 的長度，直到每種序列各有五千條為止，最後再將這些五千條 500bp 長度的序列檔案合併一起，共得了十五萬條的原核生物測試序列。

陸、 真核生物測試序列

欲知道真核生物序特徵資料庫在分類真核生物序列片段時，在不同分類階層其分類準確度為何？因此需要真核生物序列片段進行測試，而真核生物的序列皆從真核生物序列特徵資料庫中挑選，利用程式 `fungi-CandiadaA.py`(附錄十八)共挑出了三十種真核生物(表格三)的染色體序列，利用程式 `ran.py`(附錄十七)隨機切割 500bp 的長度，直到每種序列各有五千條為止，最後再將這些五千條 500bp 長度的序列檔案合併，得十五萬條的真核生物測試序列。

柒、 資料庫以外的測試序列

通常在多源性基因體學的資料當中，大多為未知的序列，因此測試我們所建構的原核生物序列特徵資料庫在分類不包含於資料庫的序列時，是否也能得到高的準確度？而不包含於資料庫的序列的選取，首先是從 NCBI 網站下載的原核生物序列中，挑

選出未被完整定序的基因體序列。然而因為這些未被完整定序的序列資料中，其 FASTA 格式的序列名稱只有 gi number 的資訊，但我們並無法直接利用 gi number 搜尋 NCBI 資料庫中這些序列所對應的分類階層，故我們會先將與 gi number 對應的 taxa id 下載後，利用程式 split-gi-id.py (附錄十九) 將每一百萬個 taxa id 寫入一個檔案，共得到 253 個檔案。並利用程式 len-check.py (附錄二十) 先算出未被完整定序的序列資料中的 gi number 相對應於那些 taxa id 的檔案後，再利用程式 m-taxasearch67-128.py、m-taxasearch128-188.py、m-taxasearch188-248.py、m-taxasearch248-254.py (附錄二十一、附錄二十二、附錄二十三、附錄二十四) 分別將未被完整定序的序列資料中的 gi number 都轉換成 taxa id，再利用程式 delrepeattaxa.py (附錄二十五) 將重複的 taxa id 去除。接下來利用程式 taxa-search.py (附錄二十六) 去 NCBI 網站上搜尋這些 taxa id 所對應的分類階層並下載，再用程式 taxa-search-1.py (附錄二十七) 將各分類階層的名稱做整理，最後利用程式 taxa-search-2.py (附錄二十八) 與原先資料庫的各分類層級比對後，即可得到各層級中不存於資料庫的序列。測試序列則是利用程式 contiglenchoos-sp.py (附錄二十九) 隨機挑選出三十個(表格三)序列，而這些序列在種的層級是不存於資料庫

中的，之後再將序列隨機切割成 500bp 的長度進行測試。此外，我們還利用程式 `contiglenchoos-ge.py`（附錄三十）隨機挑選了三十個(表格三)屬層級不存於資料庫的序列，而這些序列之後也會進一步的被隨機切割成 500bp 的長度後，再進行分類測試。

捌、 對序列特徵資料庫的分類結果進行比對

所有序列分類的結果都會與該序列正確分類階層做比對，計算出分類的準確度，然而針對不同的序列特徵資料庫，則會寫出不同的程式進行比對，如程式 `alignGenus.py`（附錄三十一）被用於比對有無質體序列的原核生物序列特徵資料庫，分類二十個不同 strain 的人體腸道細菌染色體序列的結果；程式 `taxa0423pro.py`(附錄三十二)則被用於比對無質體序列的原核生物序列特徵資料庫分類三十個不同種的人體腸道細菌染色體序列、混合序列及不存於資料庫序列的結果；程式 `taxa0407euk.py`(附錄三十三)被用於比對真核生物序列特徵資料庫分類真核生物序列和混合序列的結果。

此外，在結果的分析上也會利用 t-test 比較兩種不同操作 supervised 方式與 semi-supervised 方式所得到的分類準確度，兩者間是否有差異，並以 P 值 0.05 作為判定差異的依據。

結果

壹、 RAIphy 軟體測試

我們想知道 RAIphy 軟體所提供的原核生物序列特徵資料庫在分類其它非 RAIphy 文獻中所測試的原核生物序列時，是否也能得到如同文獻 (Nalbantoglu et al.,2011) 所宣稱的高準確度。因此我使用了二十個 (表三) 不同 strain 的人體腸道細菌染色體序列共十萬條，長度均為 500bp，去測試 RAIphy 原核生物序列特徵資料庫。已知 semi-supervised 相較於 supervised 的操作方式會加入欲測試序列中的新序列特徵後，對原先的模型進行修飾，再據修飾後的模型進行分類。因此，我們預期這兩種不同的機器學習方式，對於原核生物序列特徵資料庫分類原核生物序列的結果應該會有所不同。因此在結果的部分，會將兩種操作方式所得的分類準確度一起呈現出來。

從測試的結果中可以發現(圖一)，RAIphy 所提供的原核生物序列特徵資料庫可以將這二十個不同 strain 的人體腸道細菌染色體序列分類到其所對應的屬層級，但其分類準確度會因物種的不同而有所差異。而以分類準確度的平均值來說(圖二)，在 supervised 的操作下，屬的層級其分類準確度平均值約 43.53%；在 semi-supervised 的操作下，屬層級的平均分類準確度則是約

46.05%。將兩種操作方法相互比較的結果，發現彼此間並無顯著差異(P 值大於 0.05，圖二)。推論原因是因為所測試的序列片段，大多是資料庫中存在的序列故其序列特徵多為已知，因此在 semi-supervised 的操作方式下，其序列特徵的分類模型並不會更新太多的新特徵，使得兩種不同操作模式下的分類模型所做出的分類結果並無顯著差異。

貳、 *Prevotella* 屬序列的分類結果

在上述結果(圖一)中可發現，除了 *Prevotella* 此屬的分類準確度是零之外，其他序列皆可被分類到其所相對應的屬分類階層中。為了知道 *Prevotella* 此屬的分類準確度為何是零，我們回頭驗證 RA1phy 所提供的原核生物序列特徵資料庫，發現此資料庫中並無 *Prevotella* 這個屬和 *Prevotellaceae* 這個科存在，因此資料庫在分類時便無法將 *Prevotella* 這個屬的序列分類到其所相對應的屬層級。但其他序列因為在資料庫中皆有所對應的屬出現，因此在分類時便可以將其它序列分類至其所屬的屬層級中。故這證明了若是序列的分類層級是資料庫當中所缺乏的，該序列便無法被分類到正確的分類階層中。

雖然資料庫中缺乏 *Prevotella* 這個屬及所對應的科，但我們想

知道其在更高的分類階層的分類情況是如何？故我們檢視了 *Prevotella melaninogenica*、*Prevotella ruminicola* 和 *Prevotella denticola* 三者 在門、綱和目這三層級的分類準確度。從結果(圖三) 中發現在 supervised 的操作下，*Prevotella denticola* 在門的層級分類準確度為 16.72%、綱是 16.46%、目為 16.46%；*Prevotella melaninogenica* 在門、綱和目的分類準確度分別為 10.7%、9.72% 和 9.72%，而在 *Prevotella ruminicola* 則分別是門為 8.22%、綱為 0.75%、目為 0.75%。這三種 *Prevotella* 屬的序列其分類準確度在門的層級皆小於 20% 以下，且隨著分類階層往下其分類準確度越低。於此我們初步推論當欲分類的序列其序列特徵不存於序列資料庫時，RAIphy 便可能無法將序列據其序列特徵做出正確的分類。

而在上述結果(圖三)中，雖已知其在門、綱和目層級的分類準確度皆小於 20%，但對於剩下的 80% 的序列在各分類層級中各又被分至那些類別卻是未知的。於是我們進一步追朔結果後發現 *Prevotella denticola* 在門層級的分類結果中 52.94% 序列是被分類到 *Chlorobi*，而正確的 *Bacteroidetes* 卻只佔了 16.72%，此外 *Proteobacteria* 及 *Firmicutes* 則分別佔了 11.38% 及 10.86% (圖四、表四)。在綱的部分，*Prevotella denticola* 有 52.94% 是被分類到

Chlorobia，正確的綱 *Bacteroidia* 只佔了 16.46%，其它如 *Clostridia* 佔了 6.56%、*Deltaproteobacteria* 佔了 4.96%及 *Bacilli* 佔了 4.3%(圖五、表四)。目的部分，*Chlorobiales* 佔了 52.94%，而正確的目 *Bacteroidales* 只佔了 16.46%，其它像是 *Clostridiales* 佔了 5.8%、*Bacillales* 佔了 2.98%、*Desulfuromonadales* 佔了 2.2%及 *Methanosarcinales* 佔了 1.82%(圖六、表四)。*Prevotella denticola* 在科及屬的分類部分，雖然資料庫中缺乏正確的層級 *Prevotellaceae* 科和 *Prevotella* 屬，但從結果(圖七、表四)中可以發現到有 52.94%的序列在科的分類結果中是被分到 *Chlorobiaceae* 這個科中的，相對於與 *Prevotellaceae* 親緣關係較為相近的科 *Bacteroidaceae* 和 *Porphyromonadaceae* 卻各只分別佔了 9.54%及 6.54%。而屬的部分，屬於 *Chlorobiaceae* 科的三個屬 *Pelodictyon*、*Chlorobium* 及 *Prosthecochloris*，分別各佔了 47.64%、2.76%及 2.46%(圖八、表四)。反之與 *Prevotella* 屬較為相近的 *Bacteroides* 和 *Porphyromonas* (分別源自於 *Bacteroidaceae* 和 *Porphyromonadaceae* 科)卻只佔了 9.54%及 6.52%(圖八、表四)。

從上述各階層的分類結果中，我們可以知道測試序列 *Prevotella denticola* 雖然在門、綱、目的階層中皆有少數序列被正確的分類，但大部分的序列並未如預期的被分類至正確的分類階層中，反而

多數是被分類於門 *Chlorobi* 的支系中。在科和屬的部分也如同前面較高分類層級的結果，只有少數被分類至與其親緣關係較為相近的科和屬中，如 *Bacteroides* 和 *Porphyromonas*。故推測我們所使用的 *Prevotella denticola* 測試序列其序列特徵可能與 *Chlorobi* 支系中的 *Pelodictyon* 屬是很相近的，才會將 47.64% 的測試序列皆分類於此屬中。

至於 *Prevotella ruminicola* 測試序列在門層級的分類中 *Proteobacteria* 佔了最多有 56.16%，其次是 *Firmicutes* 佔了 12.36%，正確的門 *Bacteroidetes* 只佔了 8.22%，此外 *Euryarchaeota* 及 *Chlorobi* 這兩門則各佔了 7.28% 和 5.06% (圖九、表五)。綱的部分，*Proteobacteria* 的分支 *Gammaproteobacteria* 依舊佔了最大部分 42.62%，正確的綱 *Bacteroidia* 仍只佔了少部分 7.5%，其它如 *Firmicutes* 的分支 *Bacilli* 及 *Clostridia* 則各別佔了 6.98% 及 5.38%，此外 *Chlorobi* 門的分支 *Chlorobia* 則佔了 5.06% (圖十、表五)。

Prevotella ruminicola 在目的部分，*Alteromonadales* 佔了分類的最大部分 24.84%，*Enterobacteriales* 及 *Vibrionales* 則各佔了 5.5% 和 4.26%，而正確的目 *Bacteroidales* 依舊只佔了少部分 7.5%，其他目如 *Chlorobiales* 則佔了 5.06% (圖十一、表五)，如同前面 *Prevotella denticola* 的狀況，*Prevotella ruminicola* 的科

Prevotellaceae 和屬 *Prevotella* 亦是不存於資料庫中的。觀察 *Prevotella ruminicola* 在科分類的結果顯示(圖十二、表五)，*Shewanellaceae* 佔了 20.54%，*Enterobacteriaceae* 佔了 5.5%，*Vibrionaceae* 則佔了 4.26%，值得注意的是與 *Prevotellaceae* 親緣關係較近的科 *Bacteroidaceae* 只佔了少部分 5.06%，此外 *Chlorobiales* 的分支 *Chlorobiaceae* 則佔了 5.05%(圖十二、表五)。在屬的部分，*Shewanella* 佔了 20.54%，*Vibrio* 則佔了 3.66%，而與 *Prevotella* 親緣關係較為相近的 *Bacteroides* 屬則只佔了 5.06%(圖十三、表五)。

最後我們觀察 *Prevotella melaninogenica* 的結果，發現在門的分類階層有 43.08% 是被分類到 *Firmicutes*，27.32% 是被分到 *Proteobacteria*，正確的門 *Bacteroidetes* 只有 10.7%(圖十四、表六)。綱的部分 *Bacilli* 和 *Clostridia* 佔了分類的最大部分，分別是 27.4% 和 15.68%；其次是 *Gammaproteobacteria* 和 *Alphaproteobacteria* 分別佔了 12.94% 和 12.42%；而正確的綱 *Bacteroidetes* 只佔了 9.72%(圖十五、表六)。在目的分類結果上 *Lactobacillales* 佔了最多數 18.84%，正確的目 *Bacteroidales* 只佔了 9.72%(圖十六、表六)。科的部分，*Streptococcaceae* 和 *Anaplasmataceae* 兩個科分別佔了分類的最多數 17.36% 和 8.8%，而親緣關係與 *Prevotella melaninogenica* 較接近的科 *Bacteroidaceae* 只佔了 7.26%(圖十七、

表六)。屬的部分，*Streptococcaceae* 的分支 *Streptococcus* 佔了 17.34%，而親緣關係與 *Prevotella melaninogenica* 較接近的屬 *Bacteroides* 則佔了 7.26%(圖十八、表六)。

從上述 *Prevotella ruminicola* 和 *Prevotella melaninogenica* 的結果中，發現其序列在各分類階層中的分類並未集中的被分類在某個層級的類別內。*Prevotella ruminicola* 大部分的序列雖然都被分類在 *Proteobacteria* 的支系中，但隨著分類階層越往下，集中被分類到特定類別的比例也隨之下降，即使到了屬 *Shewanella* 也只佔了 20.54%。而 *Prevotella melaninogenica* 雖然大部分的序列是被分類於 *Firmicutes* 的支系中，但是隨著分類階層越往下，集中被分類到特定類別的比例也隨之下降，如同 *Prevotella ruminicola* 的結果，即使到了屬 *Streptococcus* 也只佔了全部序列的 17.34%。兩者相較於 *Prevotella denticola* 的分類結果，有 47.64% 都被分類到 *Pelodictyon* 屬而言，是少了很多的。故我們推論因為 RA1phy 所提供的序列資料庫中無 *Prevotella ruminicola* 和 *Prevotella melaninogenica* 的序列特徵，因此分類模型便無法將序列據其序列特徵做出明確的分類。

參、 測試序列 *Lactococcus garvieae* Lg2 的分類結果

從 RA1phy 測試的結果中(圖一)，可以觀察到 *Lactococcus garvieae* Lg2 在屬的分類結果上準確度很低，只有 7.68%，於是我們回頭檢視其在更高分類階層的分類準確度是如何？

在門的分類結果中發現，在 supervised 的操作下 *Lactococcus garvieae* Lg2 所屬的門 *Firmicutes* 在門層級的分類結果中佔了最多數 59.94%(圖十九、表七)，綱 *Bacilli* 則佔了 51.52%(圖二十、表七)，目 *Lactobacillales* 佔了 43.74%(圖二十一、表七)，科 *Streptococcaceae* 佔了 36.52%(圖二十二、表七)，在屬的部分 *Lactococcus* 佔了 7.68%(圖二十三、表七)。從 *Lactococcus garvieae* Lg2 在各層級的分類結果中，我們可以發現雖然隨著分類階層往下其分類準確度隨之下降，但 *Lactococcus garvieae* Lg2 從科到屬的分類準確度卻從 36.52%驟降到了 7.68%，而 *Streptococcus* 屬卻得到了較高的分類準確度 28.84%(圖二十三、表七)。

肆、 原核生物序列特徵資料庫測試

從 RA1phy 軟體測試中發現，其原先文獻中所建立的原核序列特徵資料庫中，所包含有被完整定序的基因體序列數目比目前 NCBI 資料庫上所更新的數目少，因此缺乏 *Prevotella* 的屬序列的

序列特徵，使其無法被正確分類。所以我們決定建構新的原核生物序列特徵資料庫對原核生物的序列片段進行分類。新的原核生物序列特徵資料庫則會分為包含質體序列和不包含質體序列兩種類型。我們預期使用這兩種資料庫來分類原核生物的序列片段可能會得到不同的結果，這是因為質體序列上的基因容易從宿主細胞中 transfer 到其他宿主細胞中，故有無加入質體序列的序列特徵可能會影響序列分類的結果。為了印證這個假說，我們利用二十個不同 strain 的人體腸道細菌染色體序列片段對有無加入質體序列的序列特徵資料庫進行測試，而操作上也分別使用 supervised 及 semi-supervised 兩種方式進行。從結果(圖二十四、圖二十五)中得知，無論是有加入質體或是未加入質體所建構的原核生物序列資料庫，在更新其內的原核生物分類階層後，皆可以將原本 RA1phy 所提供的原核序列特徵資料庫無法分類準確的 *Prevotella* 這個屬給分到對的分類層級。藉此我們得知若是欲分類的序列其分類階層皆存在資料庫當中的話，該序列是可以被準確的分類到其所屬的分類階層的。

針對這二十個不同 strain 的人體腸道細菌染色體序列的分類，從結果中得知(圖二十六)，未加入質體的原核生物序列特徵資料庫在 supervised 的操作方式下，屬的層級其分類準確度平均可

以得到 57%；而在 semi-supervised 的操作方式下，則平均可以得到 54%的分類準確度，且兩種不同的操作方法彼此間有顯著差異 (P 值小於 0.05)。此外，加入質體的原核生物序列特徵資料庫在 supervised 的操作方式下，屬的層級其分類準確度平均是 58%；而在 semi-supervised 的操作下，則是平均 57%的分類準確度，但在兩種不同的操作方法上，彼此間卻無顯著差異 (P 值大於 0.05，圖二十六)。

從這個結果中，我們可以推論出質體序列的序列特徵與其宿主細胞的染色體序列特徵可能是很雷同的，因此在有加入質體序列的資料庫在分類原核生物的序列的結果中，可以發現在 supervised 與 semi-supervised 的兩種方法的比較中並無顯著差異。

然而因為目前測試的序列中包含了同種但不同 strain 的人體腸道細菌染色體序列，為了使得測試的序列可以涵蓋更多種類的原核生物，因此將先前同種不同 strain 的物種擇一保留，其餘則置換為其它的腸道菌屬序列(置換的新序列其各個分類階層在資料庫中皆有紀錄)，總共新增了十四個新物種的序列(表三)。除此之外，因先前的測試皆只是在屬的分類準確度上有所琢磨，但為了知道在其他分類階層的分類準確度為何，故在接下來的結果中

對於這些新序列在各分類階層的分類準確度皆會有所記錄。至於為何選擇未加入質體序列的原核生物序列特徵資料庫進行測試的原因，是因為在之後真核生物序列特徵資料庫的建構上，並不會加入質體序列。從分類三十種人體腸道細菌染色體序列的結果(圖二十七)中發現，雖然隨著分類階層越來越低，其分類的準確度的平均值會降低，但在屬層級的分類平均值上仍有 68% 的準確度。此外，比較了 supervised 和 semi-supervised 兩種不同的操作方式下，在各分類階層中的分類準確度是否有所差異，發現彼此間並無顯著差異(P 值大於 0.05，表八)。證明我們所建構的原核生物序列特徵資料庫在使用上是可行的。因此類似於原核生物序列特徵資料庫的建構方式，我們推論在真核生物序列特徵資料庫的建構上或許也可以使用同樣的方法去建立。

伍、 真核生物序列特徵資料庫測試

欲知 RA1phy 是否也能將真核生物的序列片段分類到對的分類階層？因此從所建構的真核生物資料庫中，挑選出三十種真核生物的序列，切成 500bp 大小的片段後作為測試序列。從結果(圖二十八)中得知，雖然隨著分類階層越來越低，其分類的準確度也會隨之下降，但在屬的層級分類上仍有 54% 的分類準確度，而且

在 supervised 和 semi-supervised 兩種不同的操作方法上，彼此間並無顯著差異(P 值大於 0.05，表九)。此結果證明了真核生物的序列亦是可以利用序列特徵進行分類的，而且測試所使用的真核生物序列其序列特徵皆是在資料庫中有被記錄過的，因此即使是使用了 semi-supervised 的操作方法，其序列特徵的模型也不會改變太多，造成兩種操作方式在比較上無顯著差異。

陸、 混合序列測試

從上述結果已知，真核生物序列可依據其序列特徵的不同，進行分類。但若被分類的序列是混合序列時(三十種真核和三十種原核序列混合)，真核生物序列特徵資料庫是否會因為受到新加入的原核生物其序列特徵所影響，導致原先的真核生物序列片段的分類準確度有所改變？相同的做法也被使用於測試未加入質體序列的原核生物序列特徵資料庫。

從結果中得知，真核生物序列特徵資料庫在分類混合序列時，是可以將真核生物序列進行分類(圖二十九)。比較 supervised 和 semi-supervised 兩種不同的操作方法，對於真核生物序列特徵資料庫在分類混合序列時，各分類階層上的結果是否有所差異，發現除了在 kingdom 的層級上無顯著差異外(P 值大於 0.05，表十)，

其餘分類階層中皆有顯著差異(P 值小於 0.05，表十)。據 semi-supervised 的原理，對於真核序列特徵資料庫來說，原核生物的序列特徵是新的、未知的，其在原本的資料庫中並不存在，因此 semi-supervised 的模型便會將這些測試的原核生物序列其序列特徵給記憶起來，修改其原先的分類模型，故才會使得在比較各階層的分類結果中在 semi-supervised 及 supervised 兩種操作方式下，大部分的分類結果是有所顯著差異的。

更進一步地我們比較真核序列特徵資料庫在分類混和序列與單只分類真核序列這兩者的結果在各分類階層中其分類的準確度上，是否有所差異？從結果中顯示(圖三十-A)，在 supervised 的操作下，皆無顯著差異(P 值大於 0.05，表十一)；但在 semi-supervised 的操作下(圖三十-B)，分類混和序列與單只分類真核序列的結果上除了 kingdom 的分類階層上無顯著差異外，其餘階層皆有顯著不同(P 值小於 0.05，表十一)。同理，這是因為在 semi-supervised 的操作下，對於真核序列特徵資料庫來說原核生物的序列特徵是未知的，重新修飾過的 semi-supervised 模型因為新增加了原核生物的序列特徵，才會造成分類結果在 semi-supervised 及 supervised 兩種操作方式下，是有顯著差異的。

而從原核生物序列特徵資料庫分類混合序列的結果(圖三十

一)中，可發現原核生物序列是可以被正確分類的。而進一步地比較 supervised 和 semi-supervised 兩種不同的操作方法，對於原核生物序列特徵資料庫在分類混合序列上是否有所差異？從結果(表十二)中顯示不同的操作方法除了在分類階層科與種上，會造成顯著差異外(P 值小於 0.05，表十)，其餘分類階層對於不同的操作方法，皆無顯著差異。

此外也更進一步地比較，當使用混和序列與單只使用原核序列測試原核序列特徵資料庫在分類結果的準確度上，是否有所不同？從結果中顯示(圖三十二)，在 supervised 的操作方式下，分類混和序列與單只分類真核序列的結果上，只在科、屬、種上有顯著差異(P 值小於 0.05，表十三)；但在 semi-supervised 的操作方式下，分類混和序列與單只分類真核序列的結果，亦只在科、屬、種上有顯著不同(P 值小於 0.05，表十三)。

在 semi-supervised 下，原核資料庫分類混合序列的結果只有在科、屬、種有顯著差異，但在門、綱、目卻沒有，推測是因為在建構原核生物序列特徵資料庫時，可用的序列數目相較於真核較多，故以資料庫的序列特徵而言原核應該較為完整，因此即便是放入了混合序列，semi-supervised 的分類模型也不會更新太多的序列特徵，使得各分類階層的分類結果不一定有顯著差異。

柒、 非資料庫序列測試

然而在上述結果中無論是真核或是原核生物的分類，其測試序列皆是資料中已知的序列，但在實際應用的多源性基因體學資料中大多為未知的序列，故了解非資料庫序列片段，在利用未加入質體的序列特徵資料庫進行分類時，是否仍能得到高的分類準確度以及 supervised 和 semi-supervised 兩種操作方式所得到的分類結果是否有所不同，皆是很重要的事。因此我們選擇了三十種不存在於原核資料中的原核序列做測試，而分類結果的呈現方式中只顯示到屬層級的分類準確度，這是因為測試的這三十種序列在種的層級中是資料庫無的，故其在種的分類準確度必定是零，因此我們想知道這些序列至少在屬以上分類層級其準確度為何？從結果（圖三十三）中發現屬的平均準確度約是 43%。而進一步地比較 supervised 和 semi-supervised 兩種不同的操作方法，對於原核生物序列特徵資料庫在分類非資料庫序列的準確度上是否有所差異？從結果(表十四)中顯示不同的操作方法除了在分類階層門、綱與目上，有顯著差異外(P 值小於 0.05)，其餘分類階層對於不同的操作方法，皆無顯著差異。

此外，我們也測試了另外三十種不存於資料庫中的原核生物序列，與上述測試序列不同的是，這些序列的屬層級在資料庫中

是無的。測試結果(圖三十四)顯示在科的分類階層上 supervised 的分類準確度約是 26%，而 semi-supervised 則有 40%。而進一步地比較 supervised 和 semi-supervised 兩種不同的操作方法，對於原核生物序列特徵資料庫在分類不存於資料庫中的屬序列時，其分類準確度上是否有所差異？從結果(表十五)中顯示兩種不同的操作方法在所有分類階層上，皆有顯著差異(P 值小於 0.05)。

從上述結果中我們得到一個結論，在分類未知序列上，semi-supervised 及 supervised 兩種不同的操作方法在分類序列上是有明顯差異的。因此在未來多源性基因體序列的資料分析上，利用 semi-supervised 的操作方式去分類未知序列會是可行的方法。

討論

從 RAiphy 所提供的資料庫所得到的分類結果中發現，*Prevotella melaninogenica*、*Prevotella ruminicola* 和 *Prevotella denticola* 三者於科和屬的分類準確度為零，且其在較高層級的分類準確度也都小於 20%，即使是與親緣關係較相近的分類階層，所佔的百分比也很低。於此我們認為是因為三者的序列特徵皆不存於資料庫中所導致。除此之外，*Lactococcus garvieae* Lg2 在屬的分類準確度只有 7.8%，且從科到屬的分類準確度驟降了 28.72%，這說明其序列特徵在科層級的分類上相較於其他層級可能有較高的辨別能力。如果我們觀察只單用一種序列特徵去分類各層級所得到的結果時(如 RAiphy)，會發現對於某些物種而言，其分類準確度會突然從某個階層中驟降下來。對於這樣的情形，我們進一步推論序列特徵對於不同的分類層級可能是有特異性的，不同的層級可能適用於不同的序列特徵。

於先前結果中已知在 supervised 和 semi-supervised 兩種操作模式下，加入質體序列的原核序列特徵資料庫在分類二十個不同 strain 的人體腸道細菌染色體序列時並無顯著差異(P 值大於 0.05)。如先前結果所提到的，因為質體序列的序列特徵與其宿主細胞的序列特徵可能很相似，因此在使用 semi-supervised 操作時才有無顯著差異的結果。然而促使質體序列與染色體序列的序列特徵相似的原因可能為何？

我們認為有可能是在演化的過程中，質體序列長期被染色體序列馴化所導致，使其序列特徵與染色體的序列特徵差異並不大。或者是質體序列 transfer 至宿主細胞的頻率並不如我們平時所想的高。

然而，在未加入質體序列的原核序列特徵資料庫分類二十個不同 strain 的人體腸道細菌染色體序列的結果中顯示，在 supervised 和 semi-supervised 兩種操作模式下，其分類準確度是有顯著差異的(P 值小於 0.05， $P=0.02879$)。對於這個結果我們認為 P 值 0.02 其實並非有很顯著的差異和意義，因為我們在顯著差異的界定上雖是以 0.05 作為標準，但是在過去許多文獻中會以較嚴苛的數值 0.01 作為其界定的標準，因此我們推論在兩種不同的操作模式下，質體序列的有無並不會影響其分類的準確度。

參考文獻

Aguilera, A., Souza-Egipsy, V., González-Toril, E., Rendueles, O., and Amils R. (2010) Eukaryotic microbial diversity of phototrophic microbial mats in two Icelandic geothermal hot springs. *Int. Microbiol.* 13, 21-32.

Altschul, SF., Gish, W., Miller, W., Myers, EW., and Lipman, DJ. (1990) Basic local alignment search tool. *J Mol Biol.* 215, 403-10.

Brady, A., and Salzberg, S.L. (2009) Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat. Methods* 6, 673-676.

Campbell, A., Mrázek, J., Karlin, S. (1999) Genome signature comparisons among prokaryote, plasmid, and mitochondrial DNA. *Proc Natl Acad Sci U S A.* 96, 9184-9

Chor, B., Horn, D., Goldman, N., Levy, Y., Massingham, T. (2009) Genomic DNA k-mer spectra: models and modalities. *Genome Biol.* 10, R108

CompostBin is described in "Sourav Chatterji, Ichitaro Yamazaki, Zhaojun Bai and Jonathan Eisen, CompostBin: A DNA composition-based algorithm for binning environmental shotgun reads , to appear in RECOMB 2008." (unpublish)

Dinsdale, E.A., Pantos, O., Smriga, S., Edwards, R.A., Angly, F., Wegley, L., Hatay, M., Hall, D., Brown, E., Haynes, M., Krause, L., Sala, E., Sandin, S.A., Thurber, R.V., Willis, B.L., Azam, F., Knowlton, N., Rohwer, F. (2008) Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS One*. 27, e1584.

Gill, S.R., Pop, M., Deboy, R.T., Eckburg, P.B., Turnbaugh, P.J., Samuel, B.S., Gordon, J.I., Relman, D.A., Fraser-Liggett C.M., and Nelson, K.E. (2006) Metagenomic analysis of the human distal gut microbiome. *Science* 312, 1355-9.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res*. 17, 377-86.

Josse, J., Kaiser, A.D., and Kornberg, A. (1961) Enzymatic synthesis of deoxyribonucleic acid. VIII. Frequencies of nearest neighbor base sequences in deoxyribonucleic acid. *J. Biol. Chem.* 236, 864-75.

Kariin, S., and Burge, C. (1995) Dinucleotide relative abundance extremes: a genomic signature. *Trends Genet* 11, 283-90.

Krause, L., Diaz, N.N., Goesmann, A., Kelley, S., Nattkemper, T.W., Rohwer, F., Edwards, R.A., and Stoye, J. (2008) Phylogenetic classification of short environmental DNA fragments. *Nucleic Acid Res.* 36, 2230-9.

Massana, R., and Pedrós-Alio, C. (2008) Unveiling new microbial eukaryotes in the surface ocean. *Curr. Opin. Microbiol.* 11, 213-8.

McHardy, A.C., and Rigoutsos, I. (2007) What's in the mix: phylogenetic classification of metagenome sequence samples. *Curr. Opin. Microbiol.* 10, 499-503.

McHardy, A.C., Martín, H.G., Tsirigos, A., Hugenholtz, P., and Rigoutsos, I. (2007) Accurate phylogenetic classification of variable-length DNA fragments. *Nat. Methods* 4, 63-72.

Nalbantoglu, O.U., Way, S.F., Hinrichs, S.H., and Sayood, K. (2011) RA1phy: phylogenetic classification of metagenomics samples using iterative refinement of relative abundance index profiles. *BMC Bioinformatics* 12, 41.

Moon-van der Staay, S.Y., De Wachter, R., and Vaulot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* 409, 607-10.

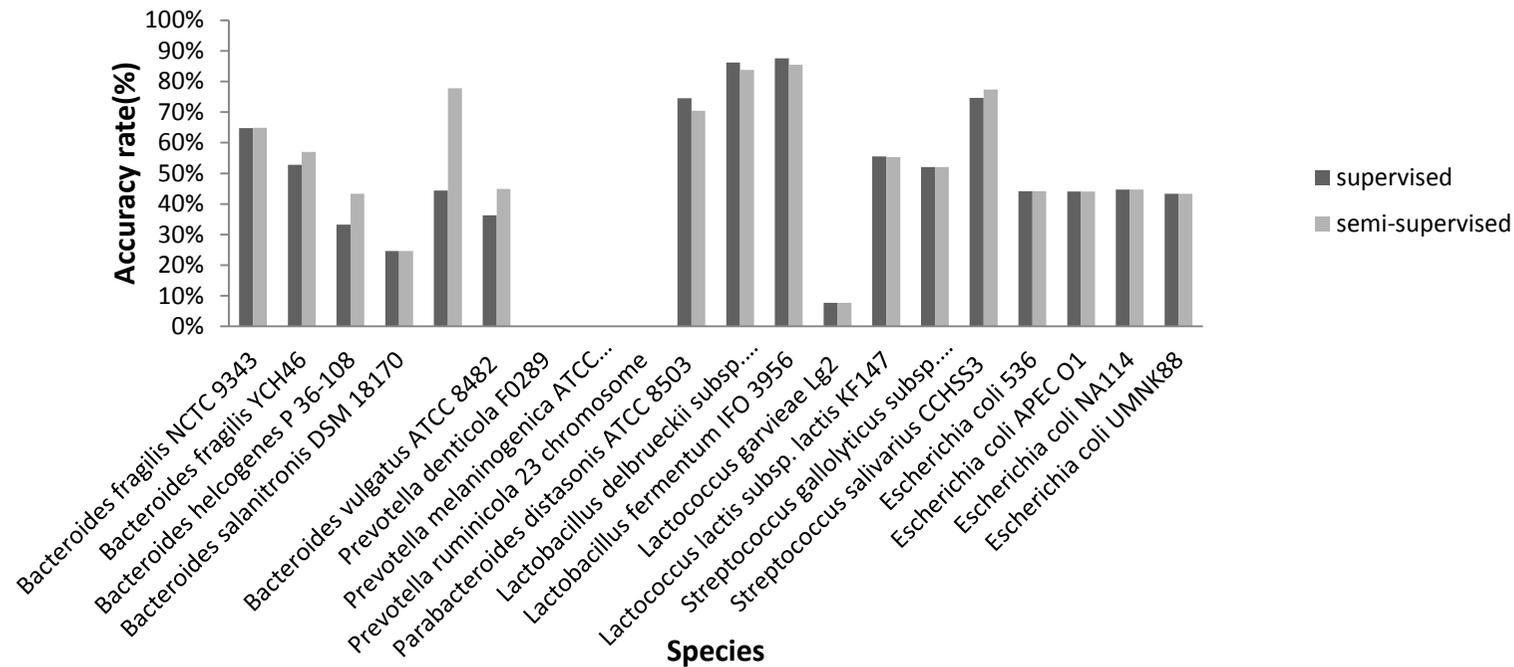
Pruitt, K.D., Tatusova, T., Klimke, W., and Maglott, D.R. (2009) NCBI Reference Sequences: current status, policy and new initiatives. *Nucleic Acids Res* 37, D32-6.

Richards, T.A., and Bass, D. (2005) Molecular screening of free-living microbial eukaryotes: diversity and distribution using a meta-analysis. *Curr. Opin. Microbiol.* 8, 240-252.

Tringe, S.G., and Rubin, E.M. (2005) Metagenomics: DNA sequencing of environmental samples. *Nat Rev Genet.* 6, 805-14.

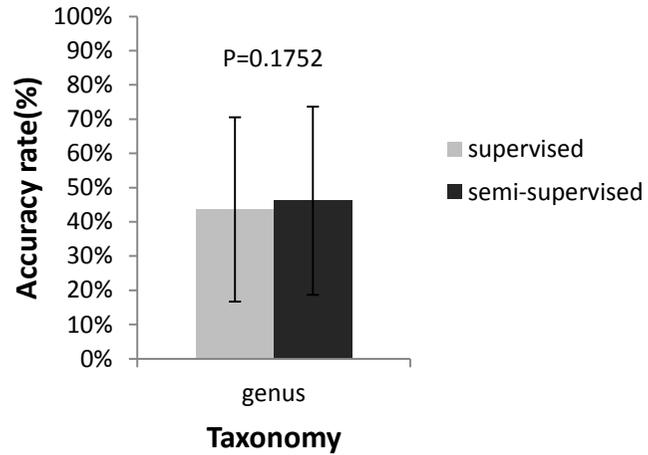
Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R., and Gordon, J.I. (2009) The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med.* 11, 6ra14.

Wooley, J.C., Godzik, A., and Friedberg, I. (2010) A primer on metagenomics. *PLoS Comput. Biol.* 6, 10-1371.



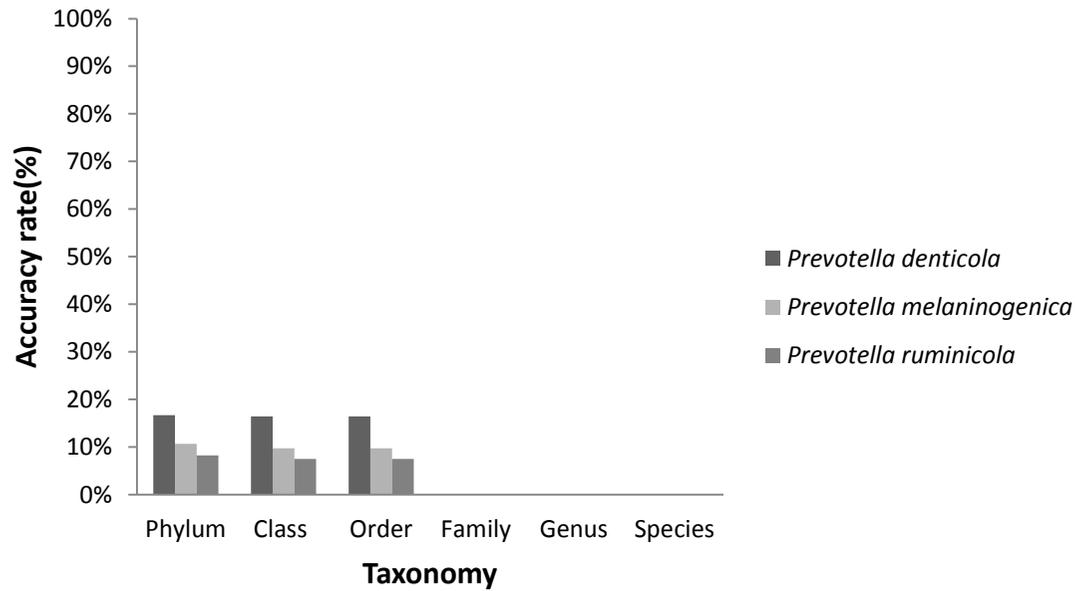
圖一、RAIphy 測試-各物種在屬層級的分類準確度

利用兩種不同的操作方法，supervised 和 semi-supervised 去測試 RAIphy 所提供的原核生物序列特徵資料庫分類二十種不同 strain、500bp 片段大小的腸胃道屬原核生物序列的結果。圖的橫軸為二十種不同 strain 的腸胃道細菌序列的名稱，縱軸的分類準確度為各原核生物序列被準確分類到屬的數目除以該原核生生物測試序列的總數。



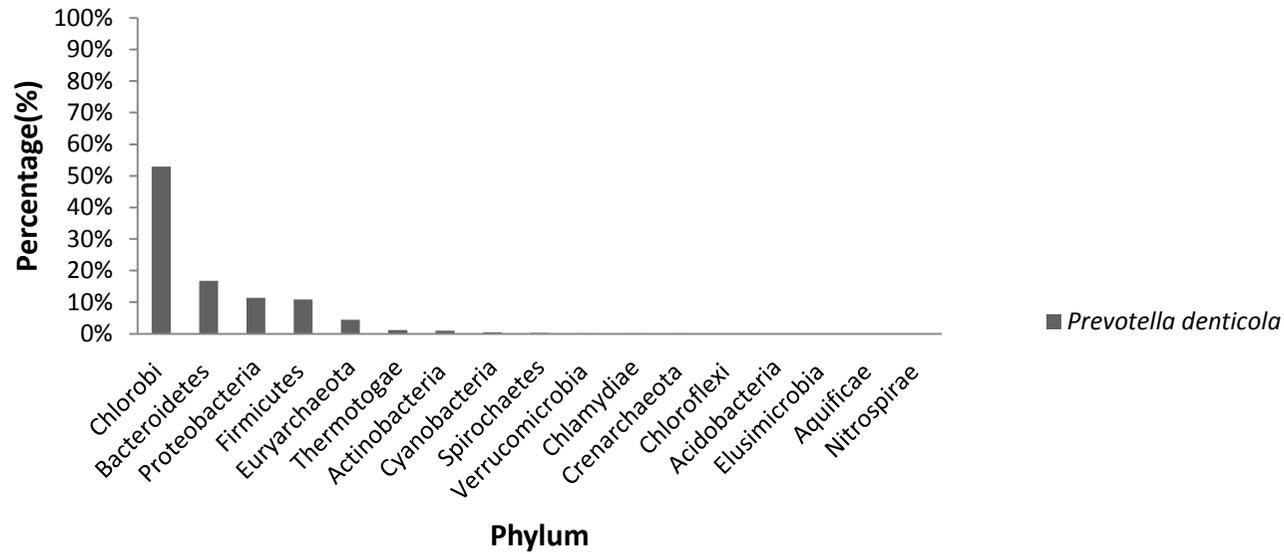
圖二、RAIphy 測試-屬層級分類的平均準確度

圖的橫軸為屬分類階層，縱軸為 RAIphy 所提供的原核生物序列特徵資料庫分類二十種不同 strain 腸胃道細菌序列的結果，其分類的平均準確度。P 值為 t-test 的比較結果，比較在 supervised 和 semi-supervised 兩種不同的操作方法下，腸胃道細菌序列的分類準確度於屬層級上的是否有所差異，結果顯示 P 值大於 0.05，故無顯著差異。



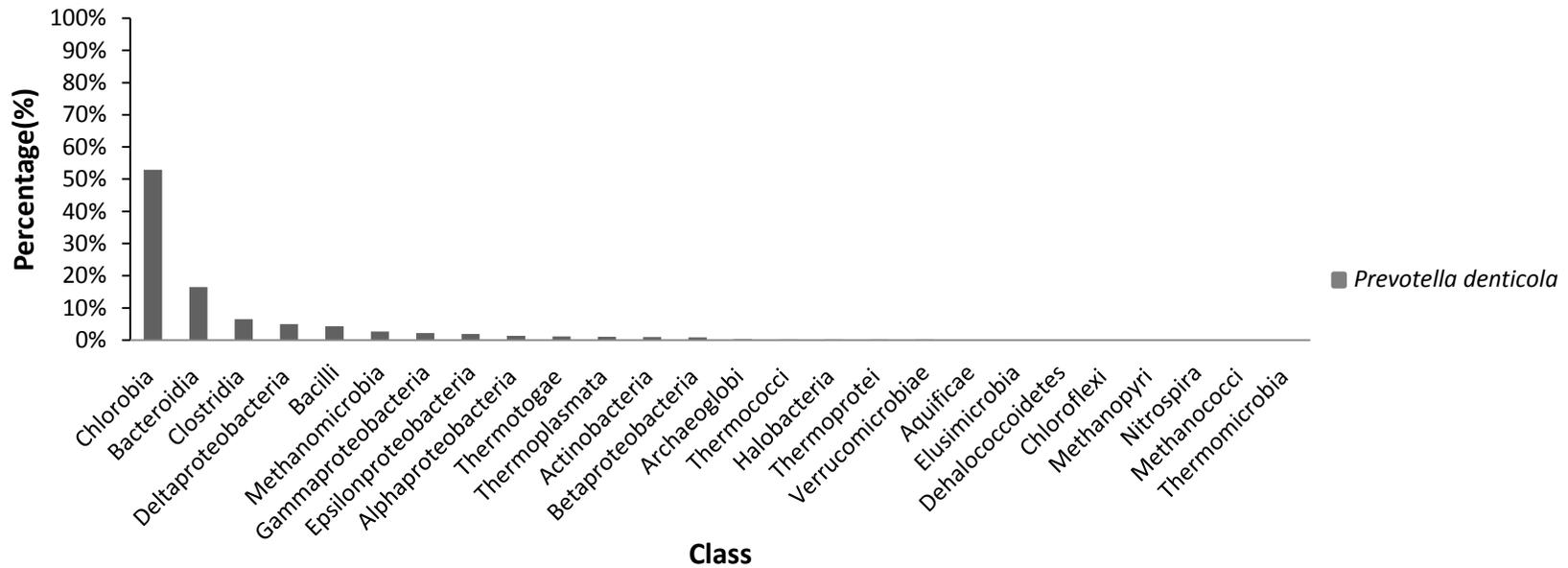
圖三、*Prevotella* 屬的三個物種的在各層級的分類結果

圖的橫軸為不同的分類階層；縱軸為分類準確度，分類準確度算法是利用該物種被分類到正確的屬的序列數目除以該物種測試序列的總數。*Prevotella* 屬的三個物種在門、綱目的分類準確皆小於 20%，在科、屬、種則為 0%。



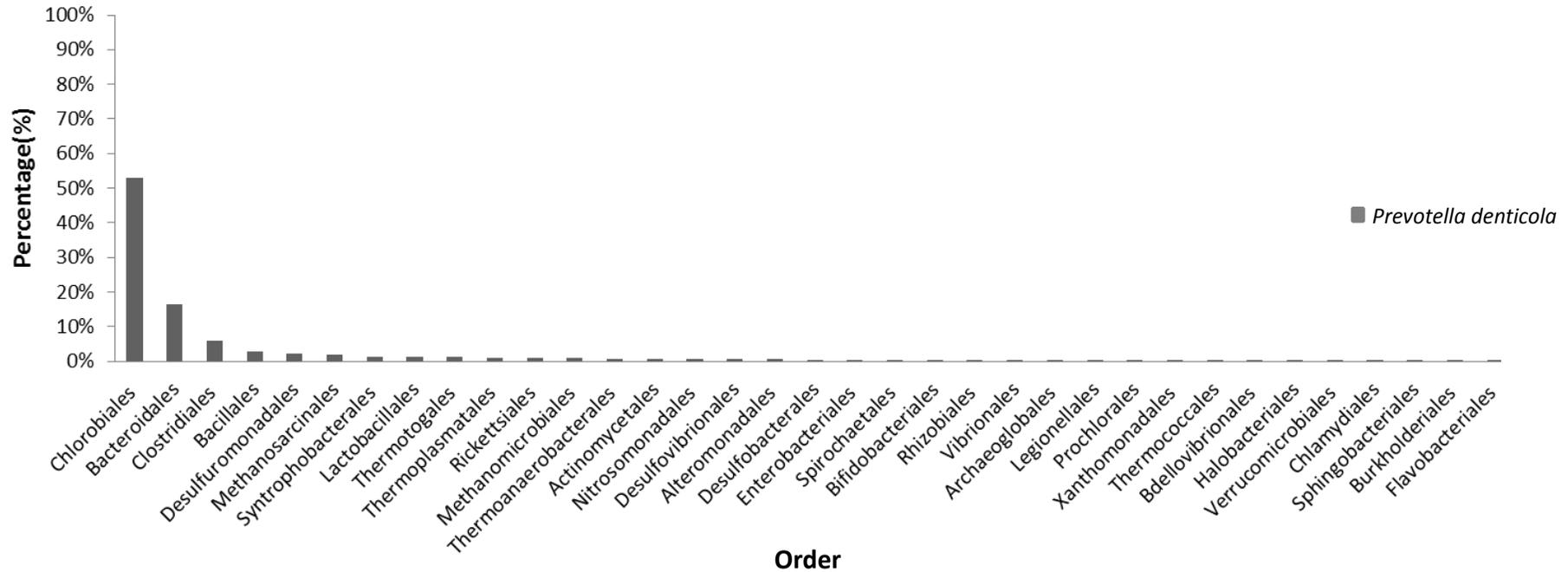
圖四、*Prevotella denticola* 在門層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在門層級的分類上各別被分至那些門中，縱軸顯示這些被分類到的門各別佔了多少的百分比。百分比的算法是利用被分類到該門的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



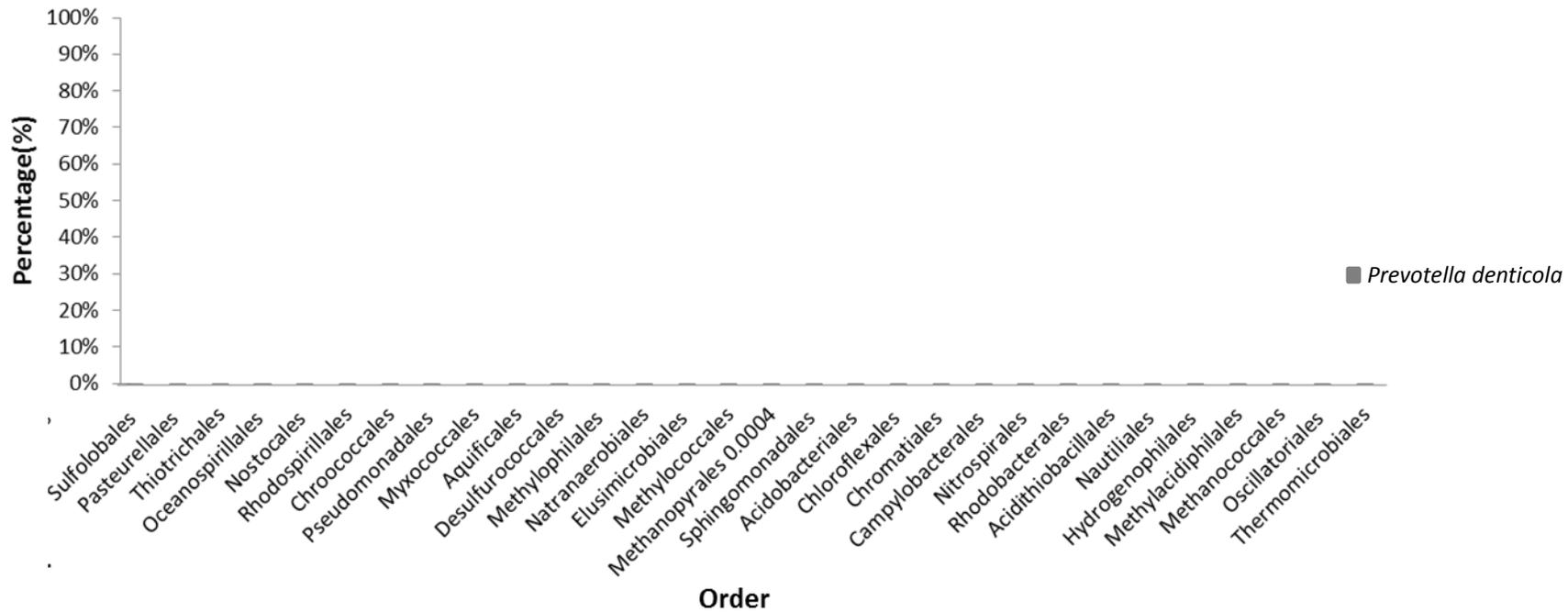
圖五、*Prevotella denticola* 在綱層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在綱層級的分類上各別被分至那些綱中，縱軸顯示這些被分類到的綱各別佔了多少的百分比。百分比的算法是利用被分類到該綱的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



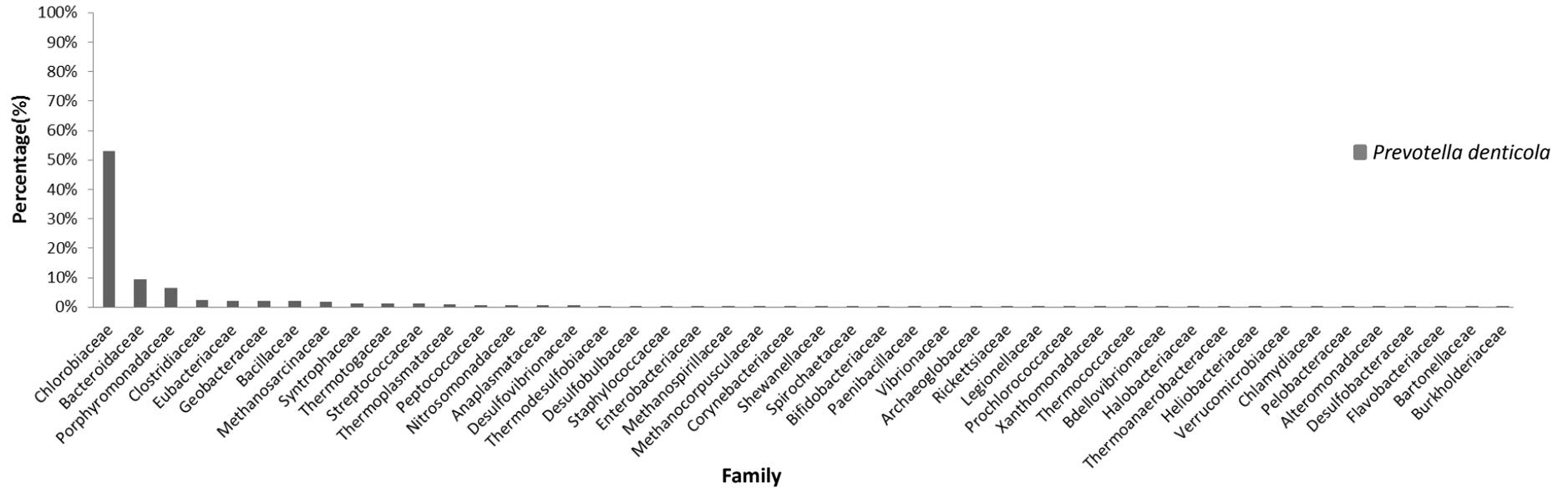
圖六、*Prevotella denticola* 在目層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在目層級的分類上各別被分至那些目中，縱軸顯示這些被分類到的目各別佔了多少的百分比。百分比的算法是利用被分類到該目的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



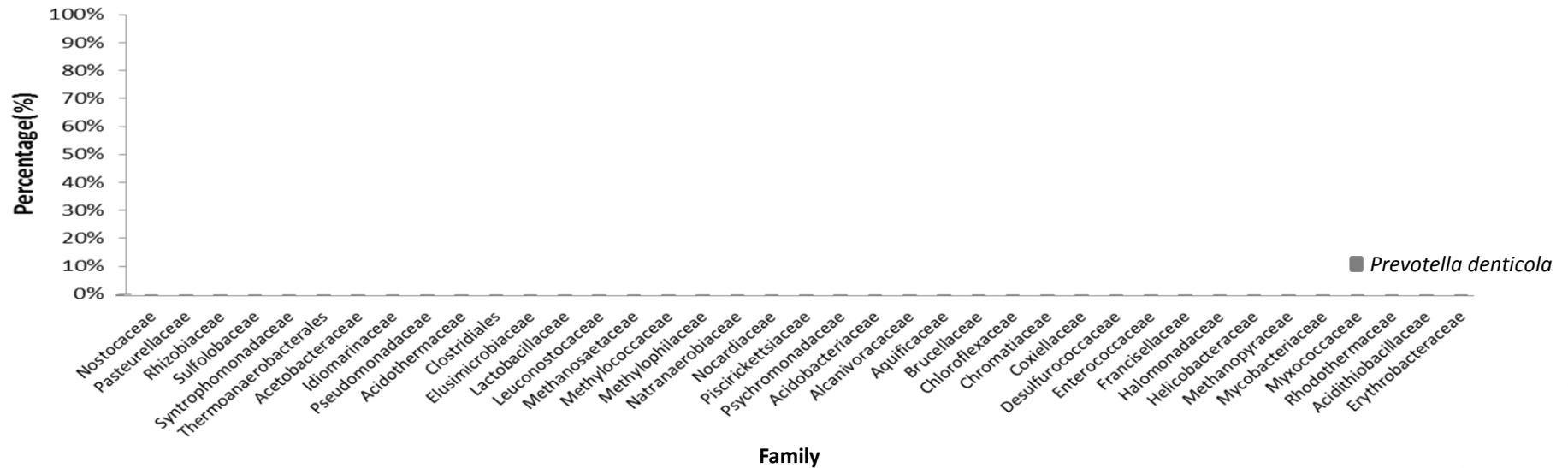
圖六、*Prevotella denticola* 在目層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在目層級的分類上各別被分至那些目中，縱軸顯示這些被分類到的目各別佔了多少的百分比。百分比的算法是利用被分類到該目的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



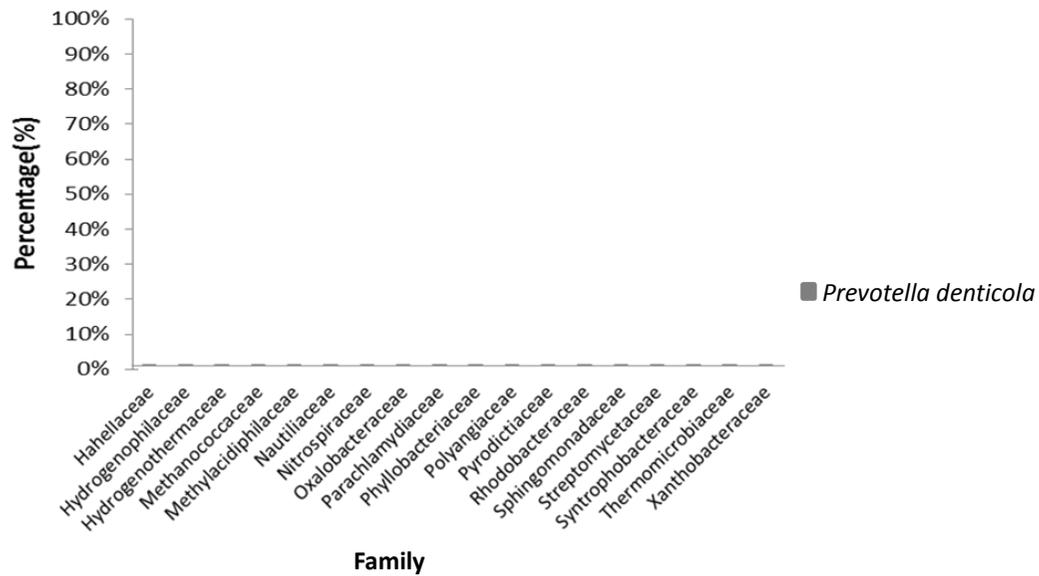
圖七、*Prevotella denticola* 在科層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



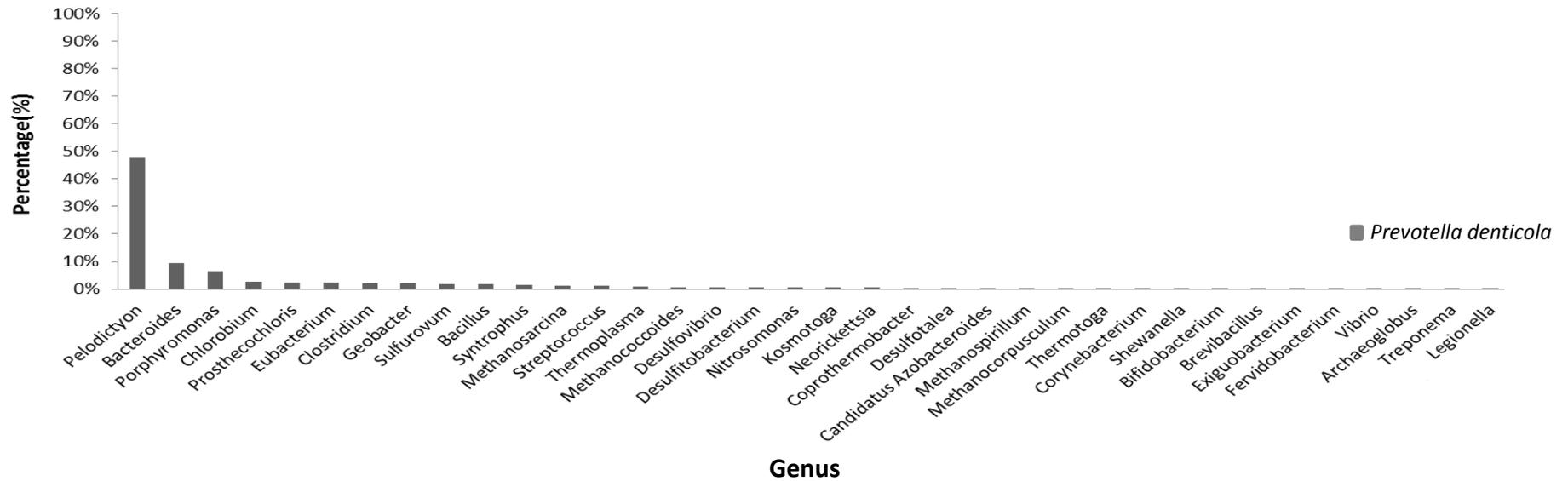
圖七、*Prevotella denticola* 在科層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



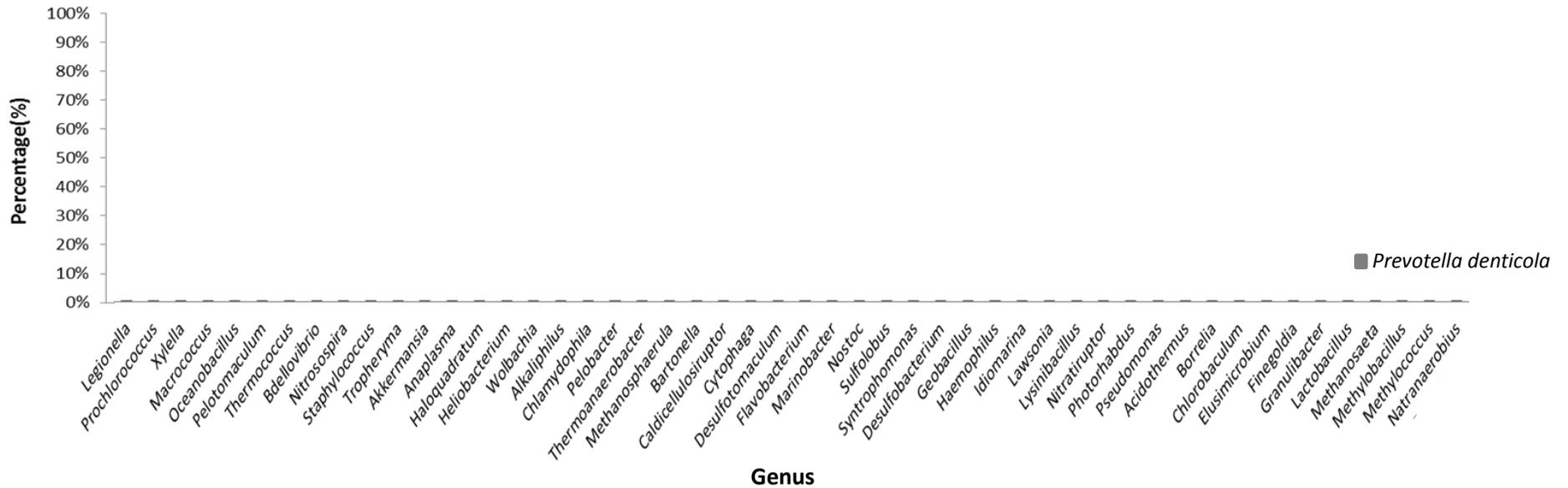
圖七、*Prevotella denticola* 在科層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



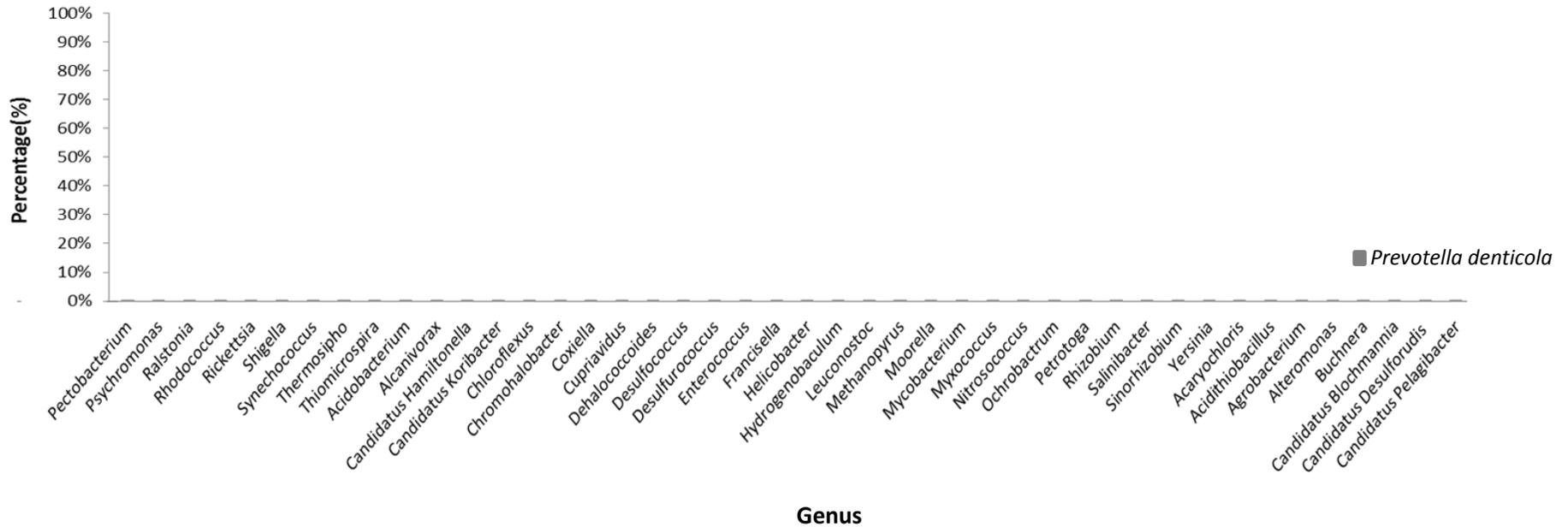
圖八、*Prevotella denticola* 在屬層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



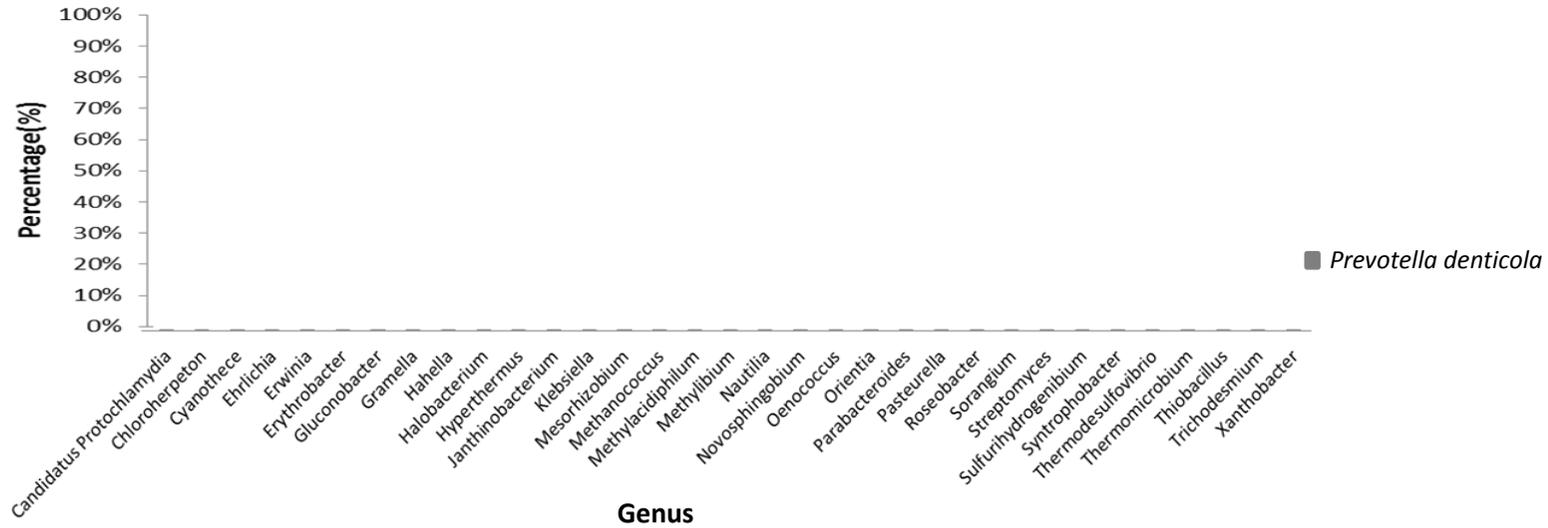
圖八、*Prevotella denticola* 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



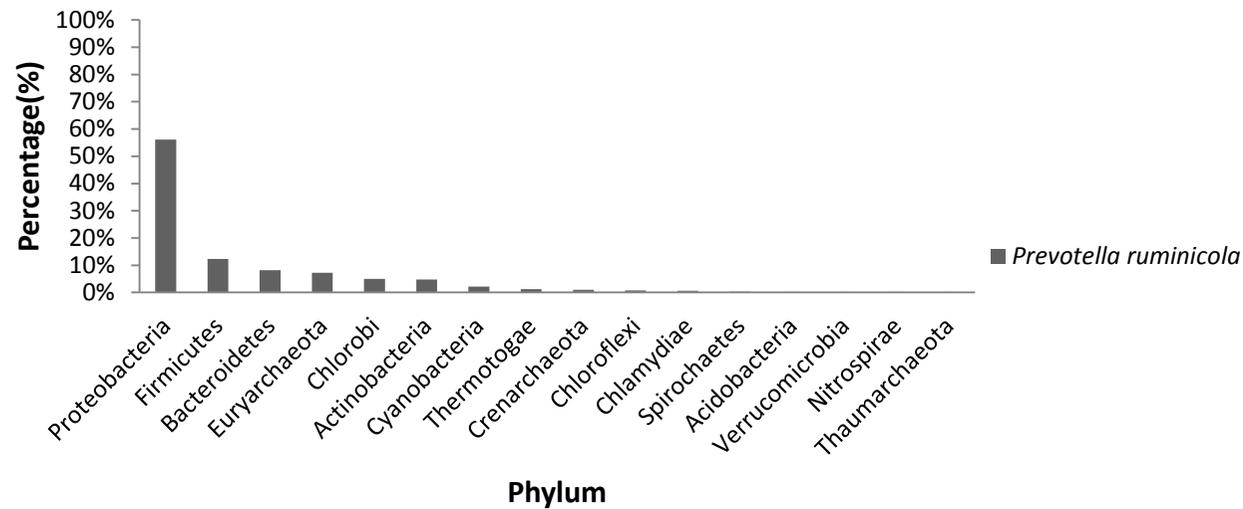
圖八、*Prevotella denticola* 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



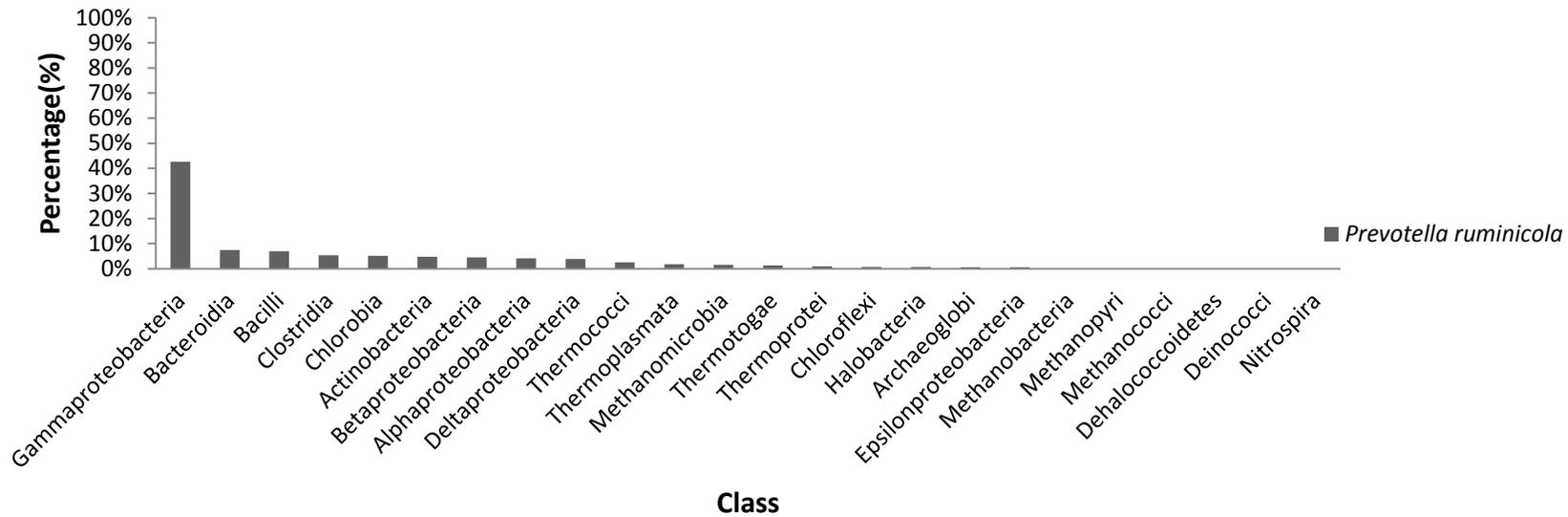
圖八、*Prevotella denticola* 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella denticola* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella denticola* 序列數目除以 *Prevotella denticola* 測試序列的總數。



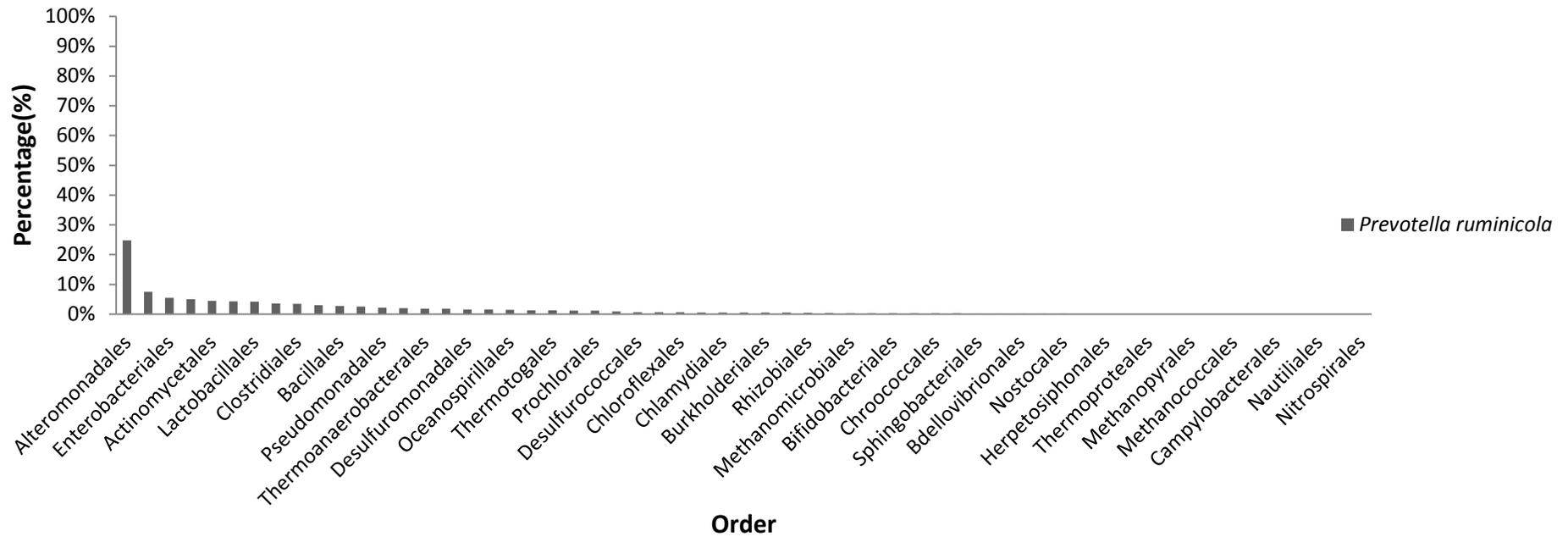
圖九、*Prevotella ruminicola* 在門層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在門層級的分類上各別被分至那些門中，縱軸顯示這些被分類到的門各別佔了多少的百分比。百分比的算法是利用被分類到該門的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



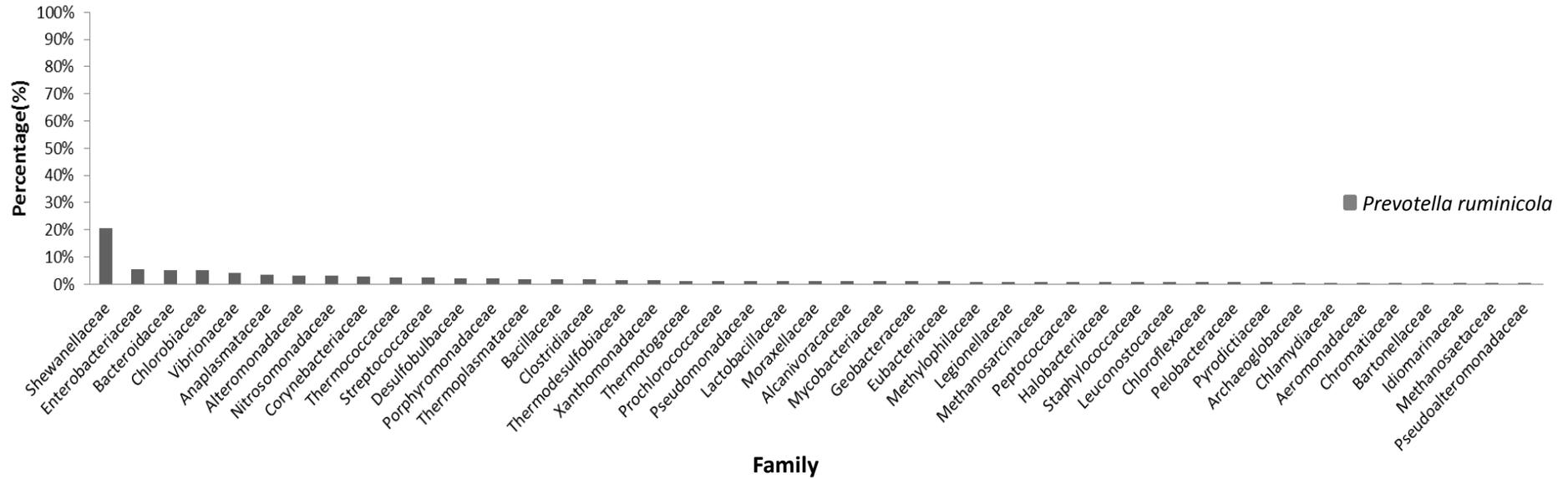
圖十、*Prevotella ruminicola* 在綱層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在綱層級的分類上各別被分至那些綱中，縱軸顯示這些被分類到的綱各別佔了多少的百分比。百分比的算法是利用被分類到該綱的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



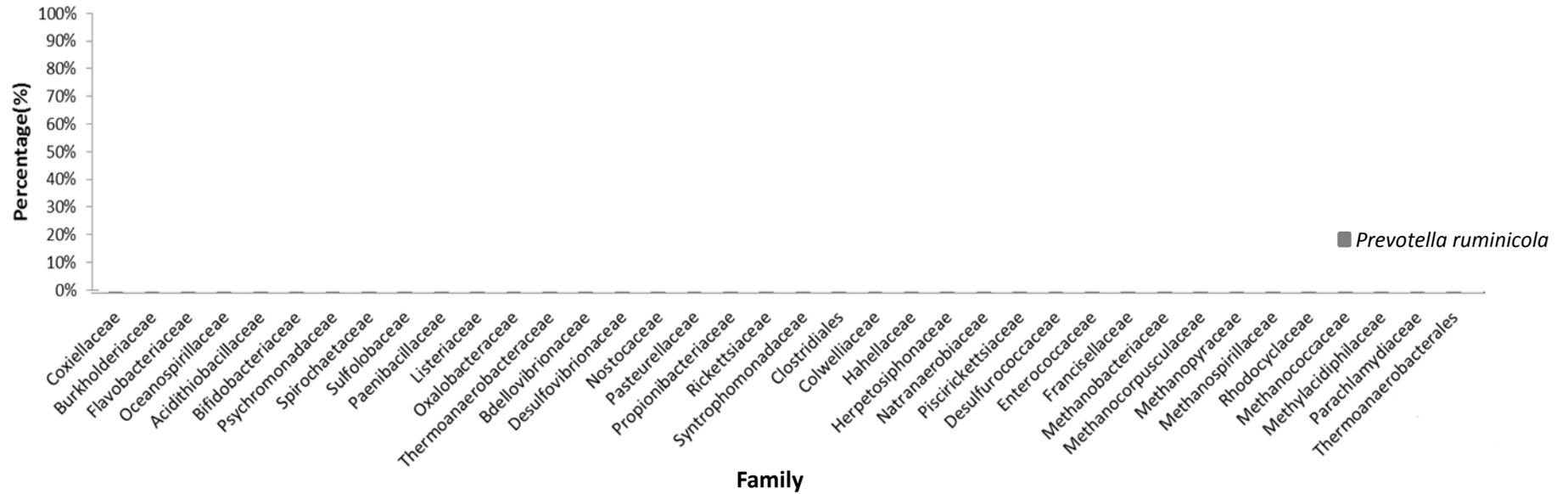
圖十一、*Prevotella ruminicola* 在目層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在目層級的分類上各別被分至那些目中，縱軸顯示這些被分類到的目各別佔了多少的百分比。百分比的算法是利用被分類到該目的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



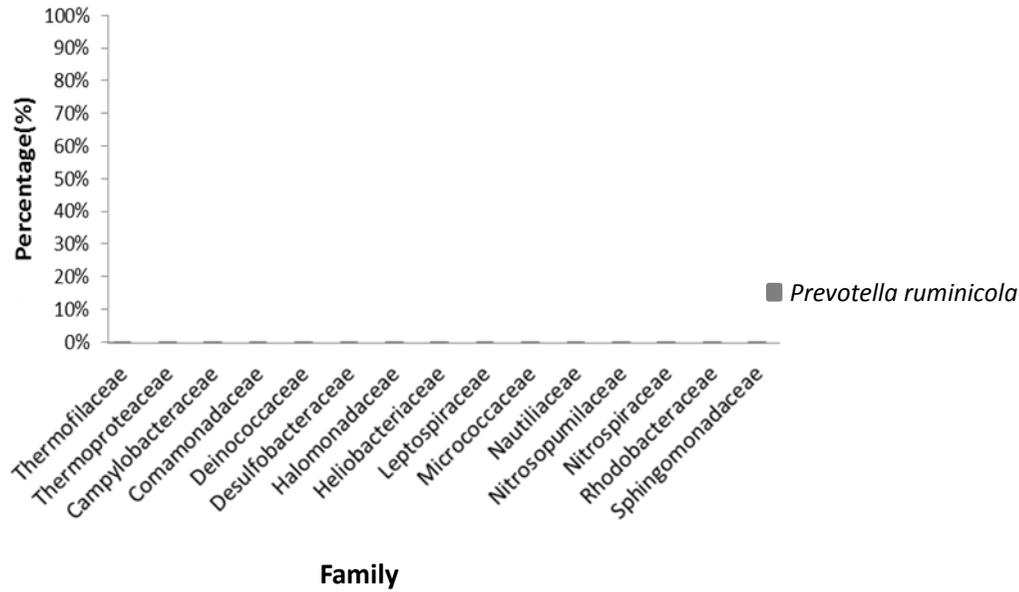
圖十二、*Prevotella ruminicola* 在科層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



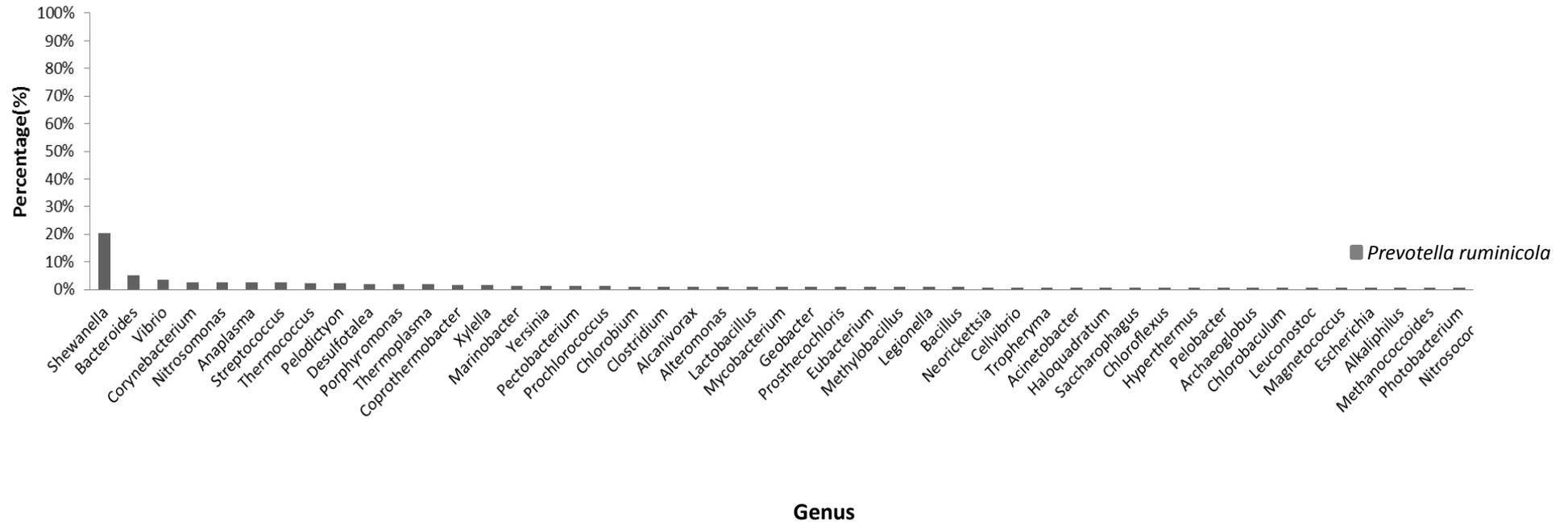
圖十二、*Prevotella ruminicola* 在科層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



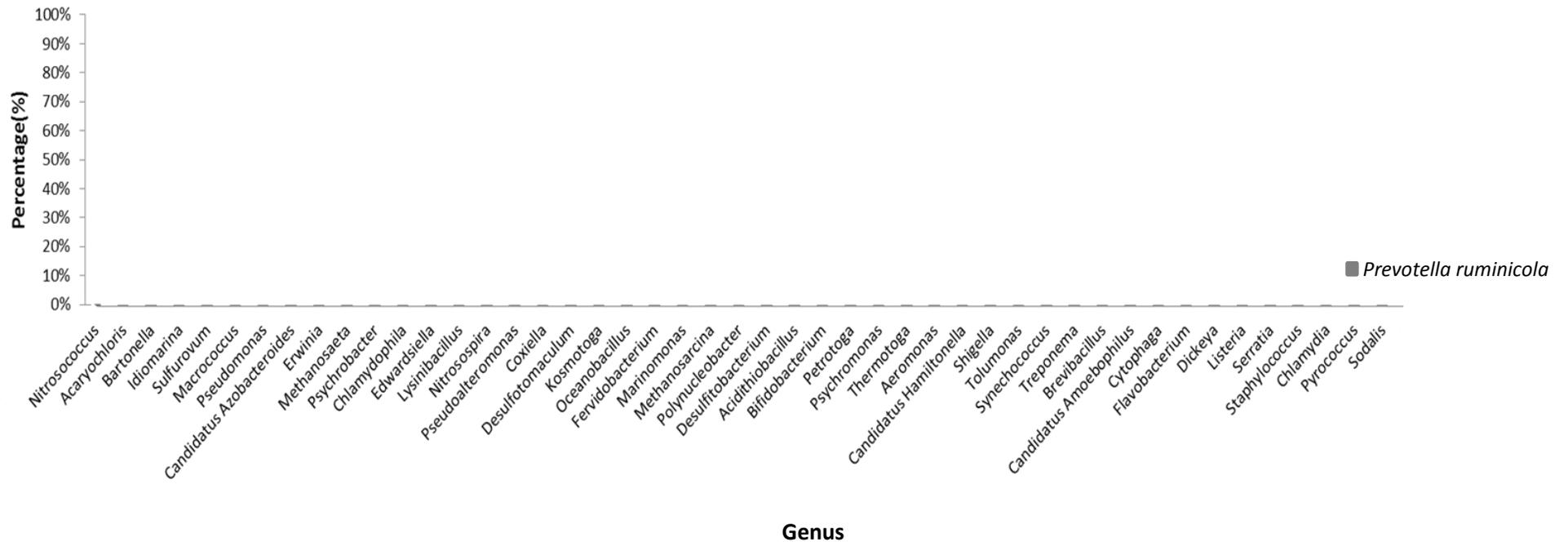
圖十二、*Prevotella ruminicola* 在科層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



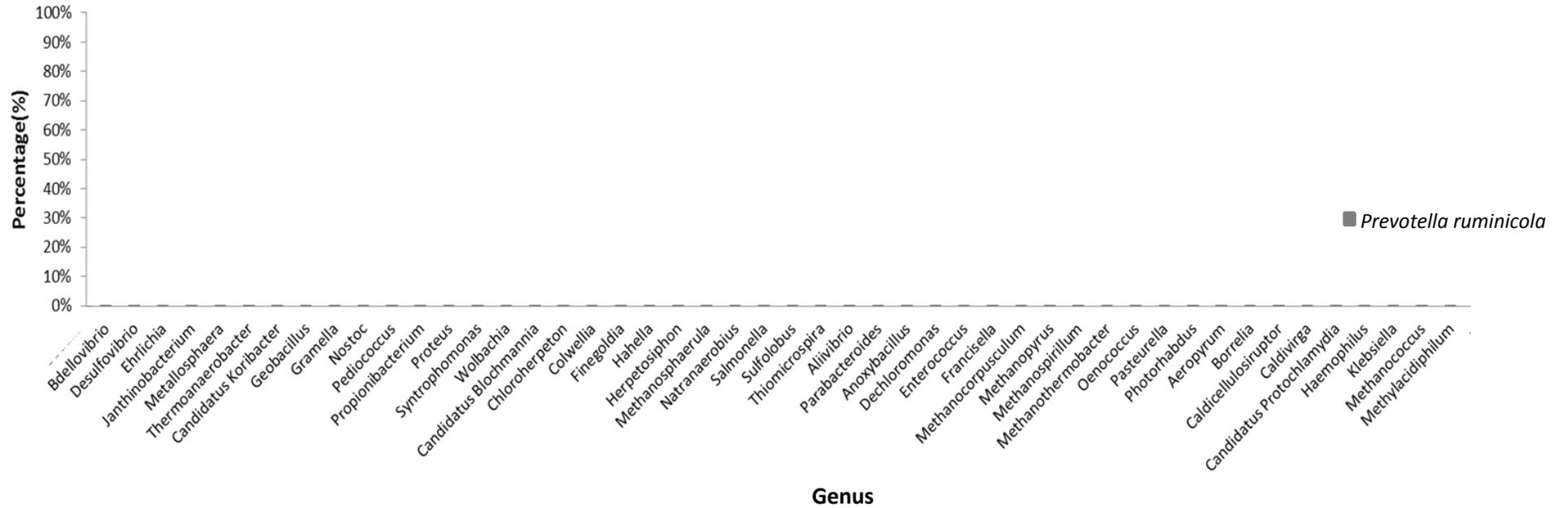
圖十三、*Prevotella ruminicola* 在屬層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



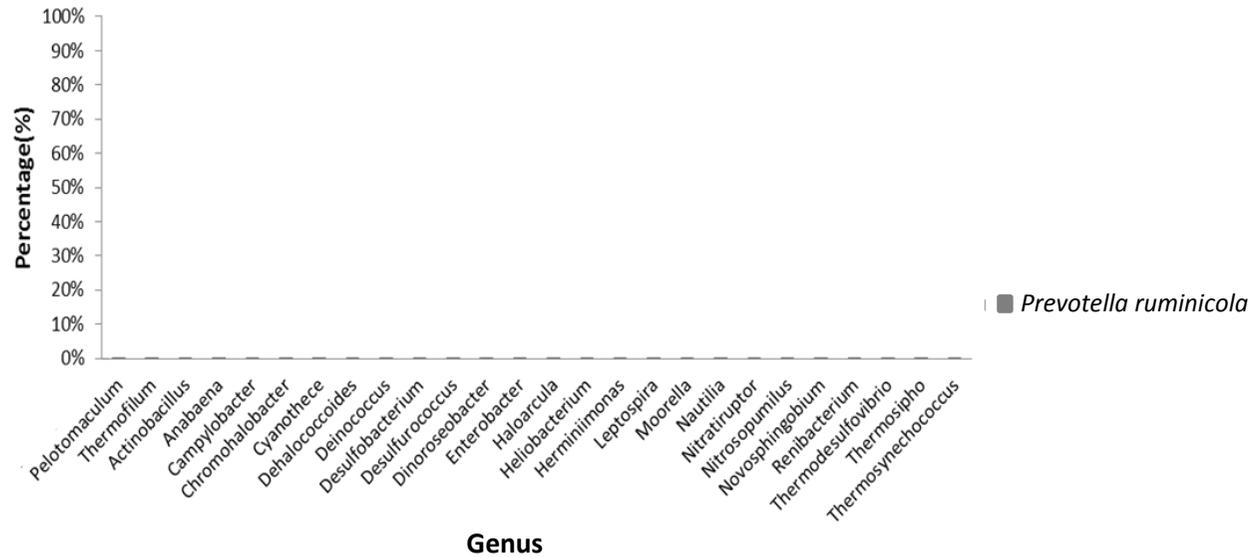
圖十三、*Prevotella ruminicola* 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



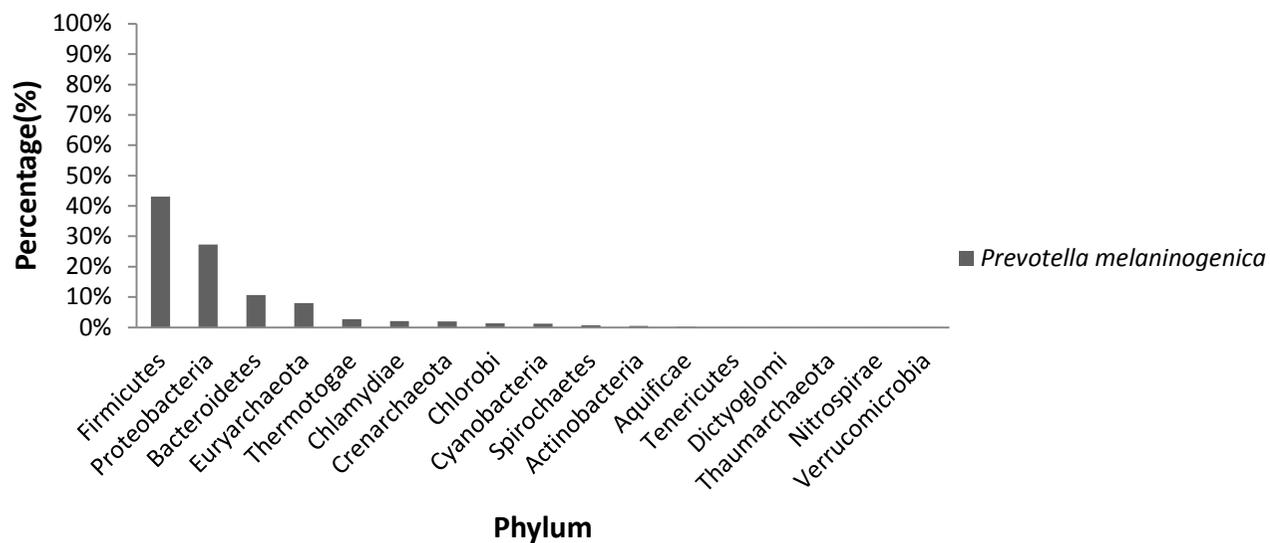
圖十三、*Prevotella ruminicola* 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



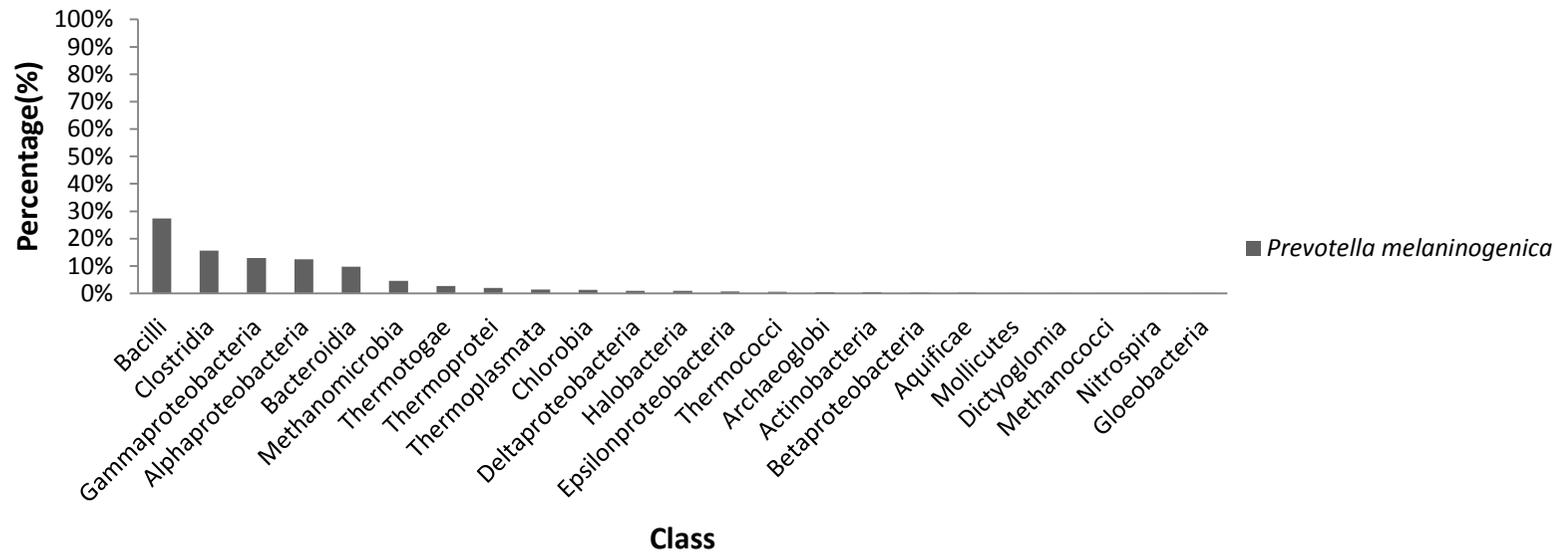
圖十三、*Prevotella ruminicola* 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella ruminicola* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella ruminicola* 序列數目除以 *Prevotella ruminicola* 測試序列的總數。



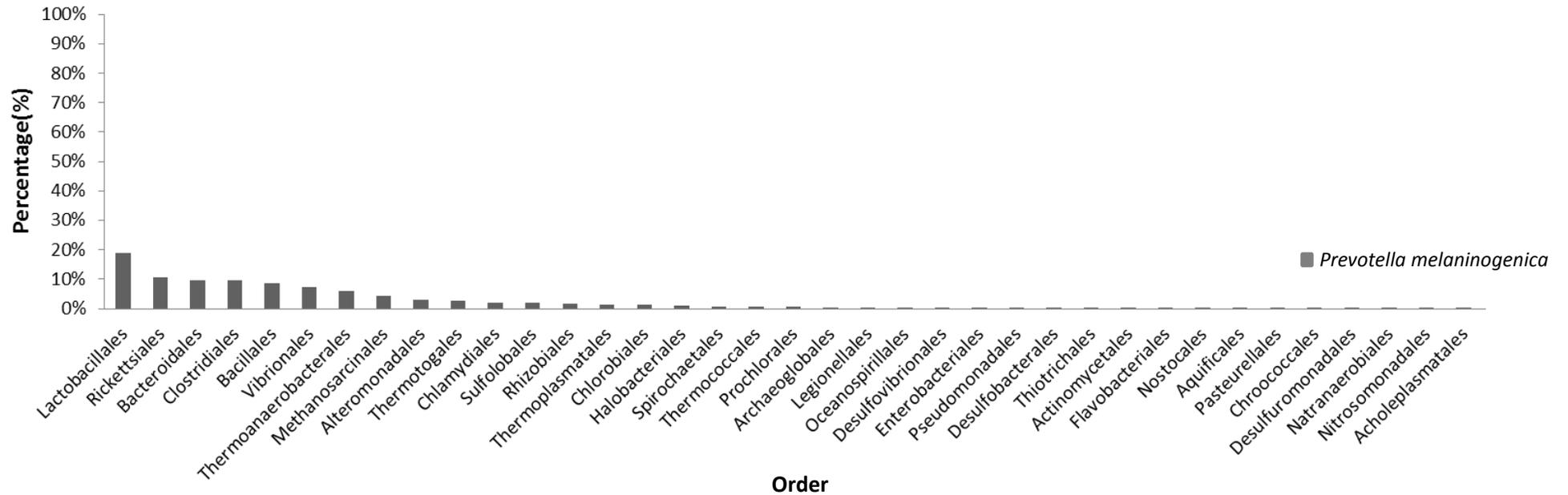
圖十四、*Prevotella melaninogenica* 在門層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在門層級的分類上各別被分至那些門中，縱軸顯示這些被分類到的門各別佔了多少的百分比。百分比的算法是利用被分類到該門的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



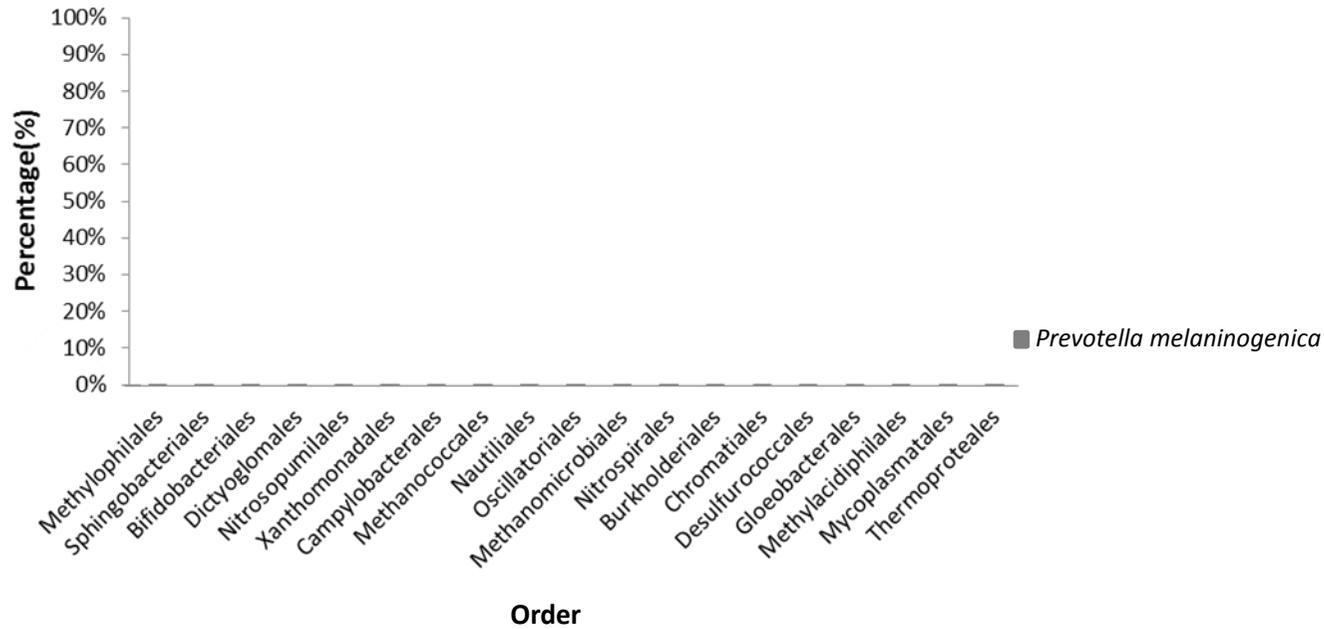
圖十五、*Prevotella melaninogenica* 在綱層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在綱層級的分類上各別被分至那些綱中，縱軸顯示這些被分類到的綱各別佔了多少的百分比。百分比的算法是利用被分類到該綱的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



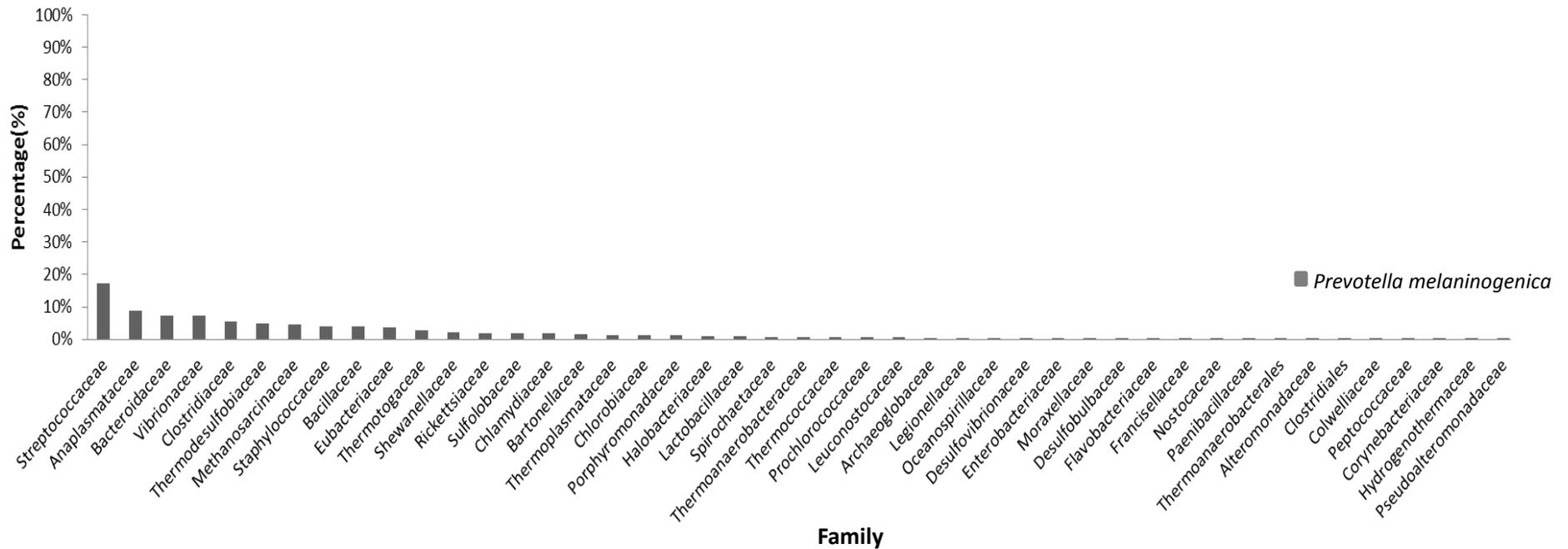
圖十六、*Prevotella melaninogenica* 在目層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在目層級的分類上各別被分至那些目中，縱軸顯示這些被分類到的目各別佔了多少的百分比。百分比的算法是利用被分類到該目的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



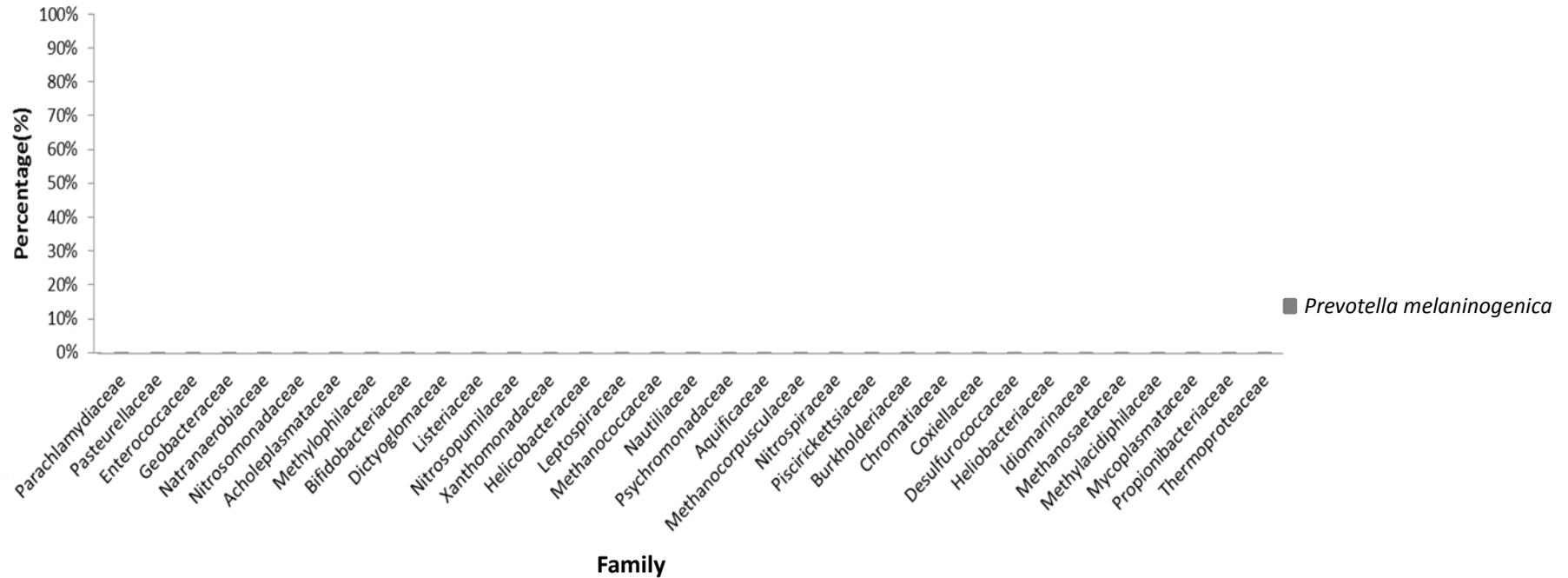
圖十六、*Prevotella melaninogenica* 在目層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在目層級的分類上各別被分至那些目中，縱軸顯示這些被分類到的目各別佔了多少的百分比。百分比的算法是利用被分類到該目的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



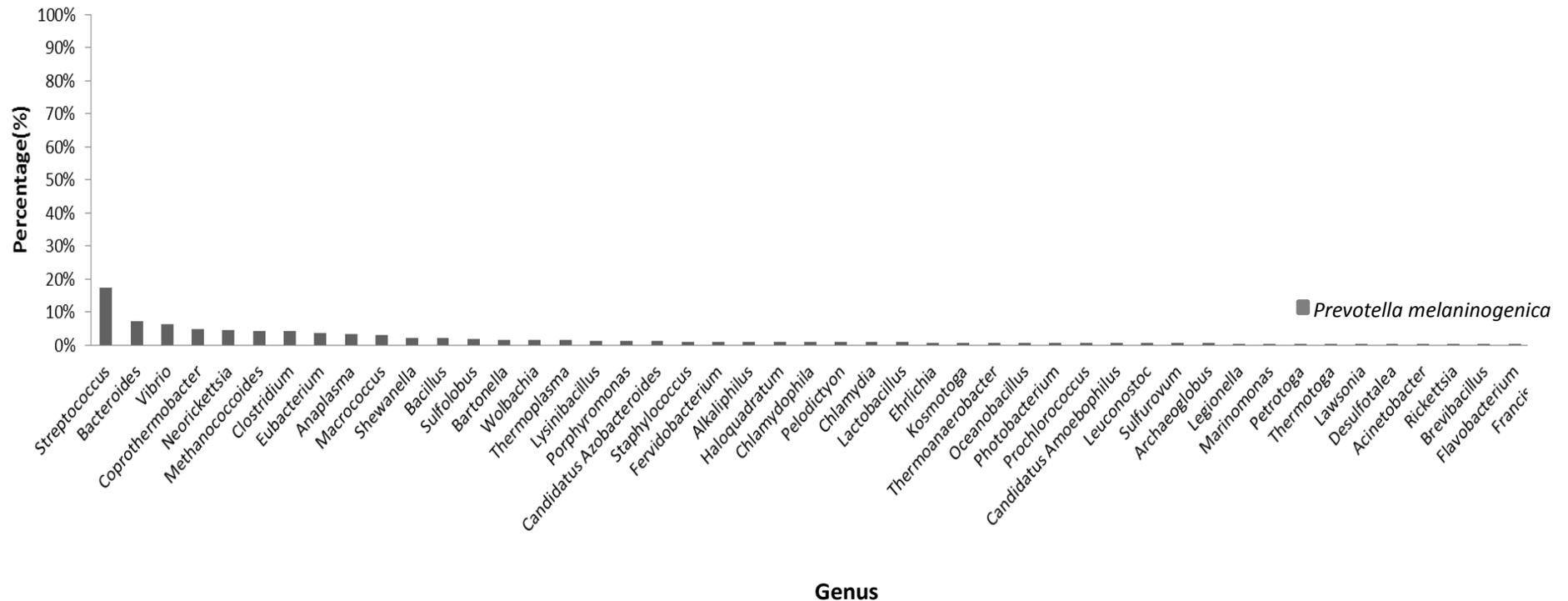
圖十七、*Prevotella melaninogenica* 在科層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



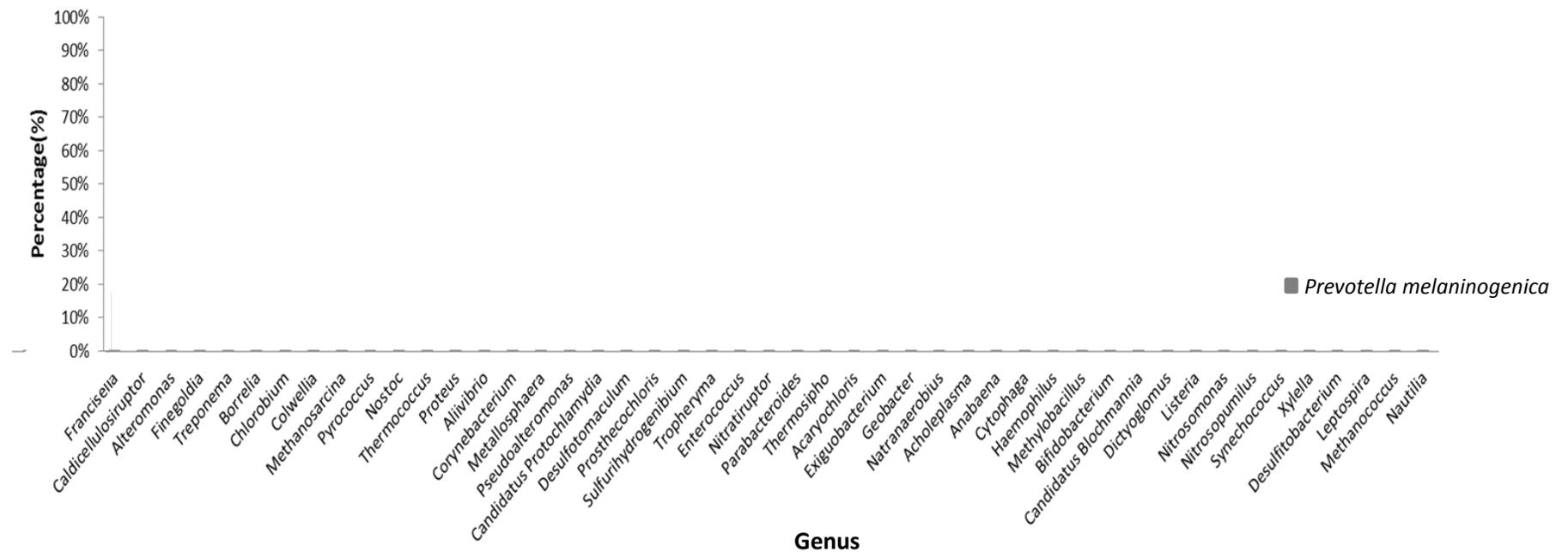
圖十七、*Prevotella melaninogenica* 在科層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



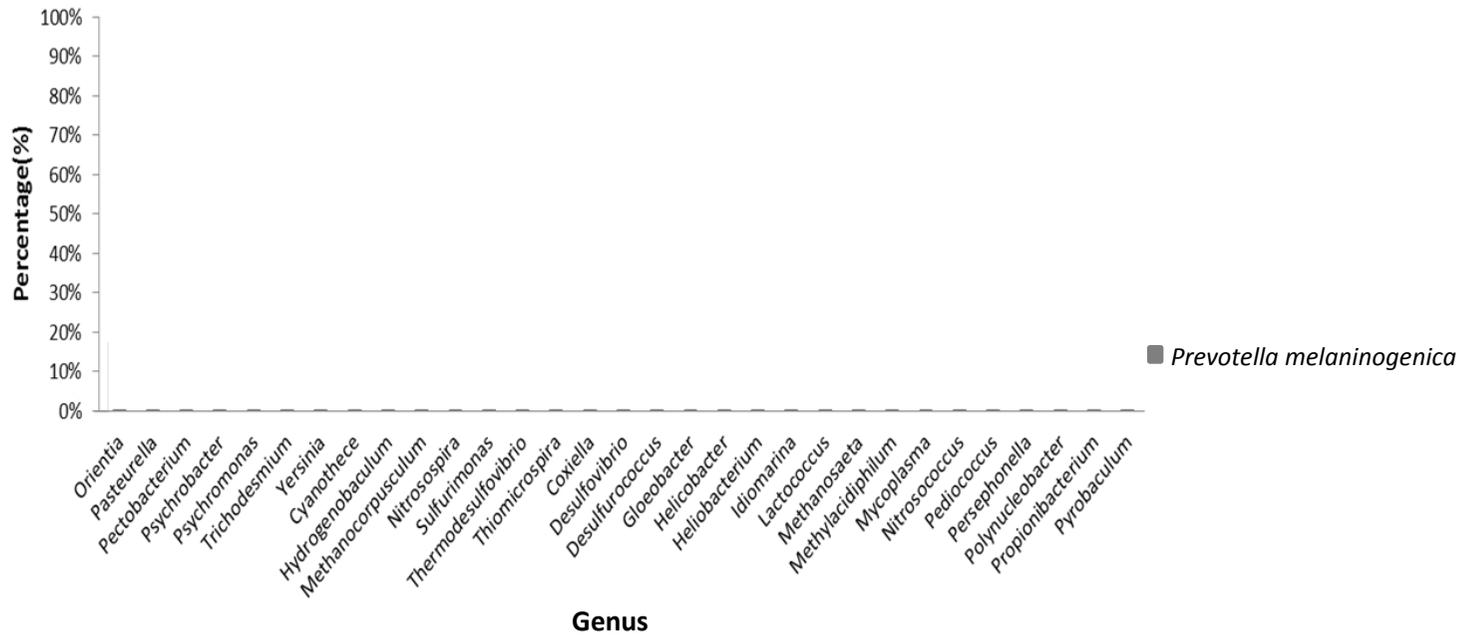
圖十八、*Prevotella melaninogenica* 在屬層級的分類結果

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



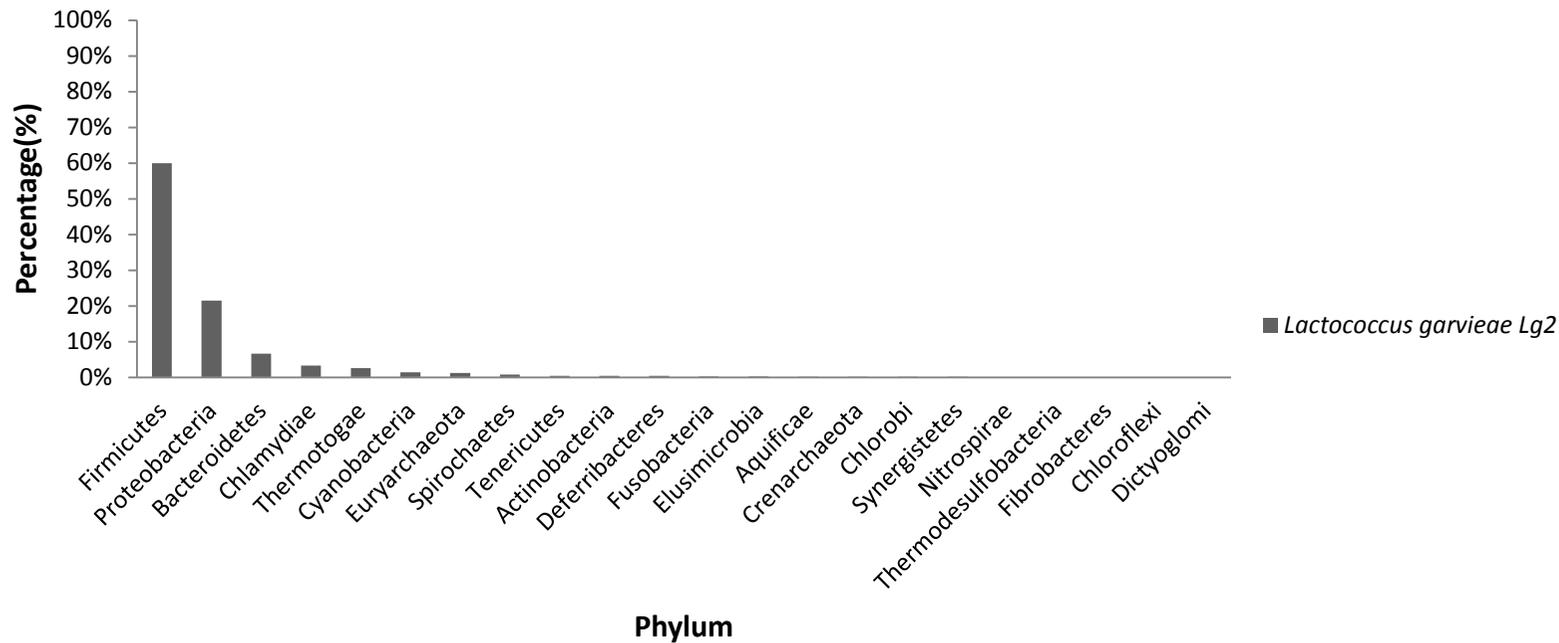
圖十八、*Prevotella melaninogenica* 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



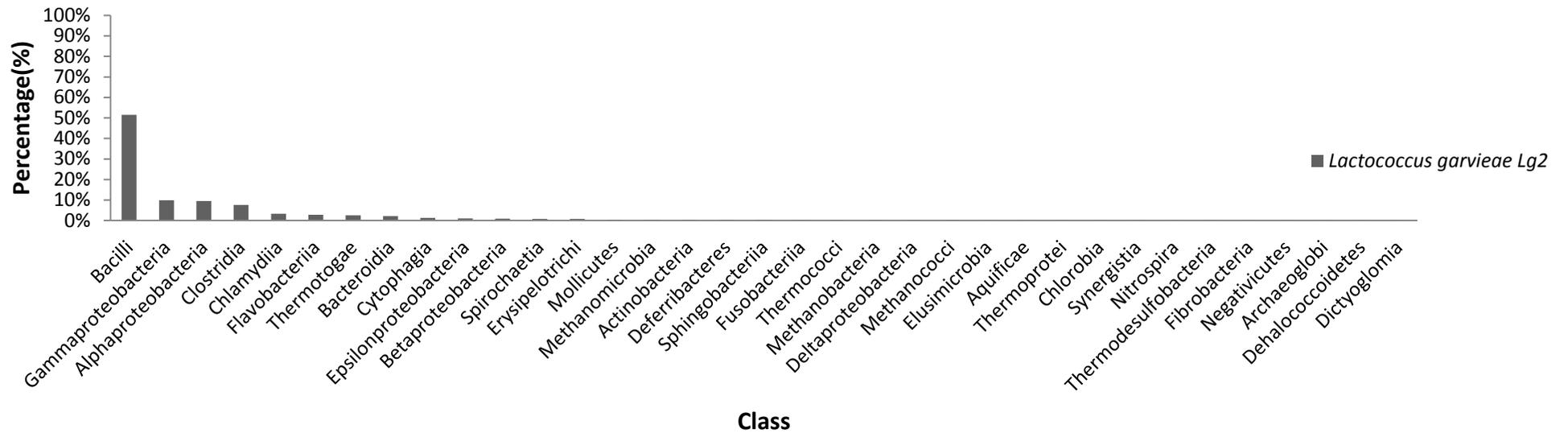
圖十八、*Prevotella melaninogenica* 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Prevotella melaninogenica* 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Prevotella melaninogenica* 序列數目除以 *Prevotella melaninogenica* 測試序列的總數。



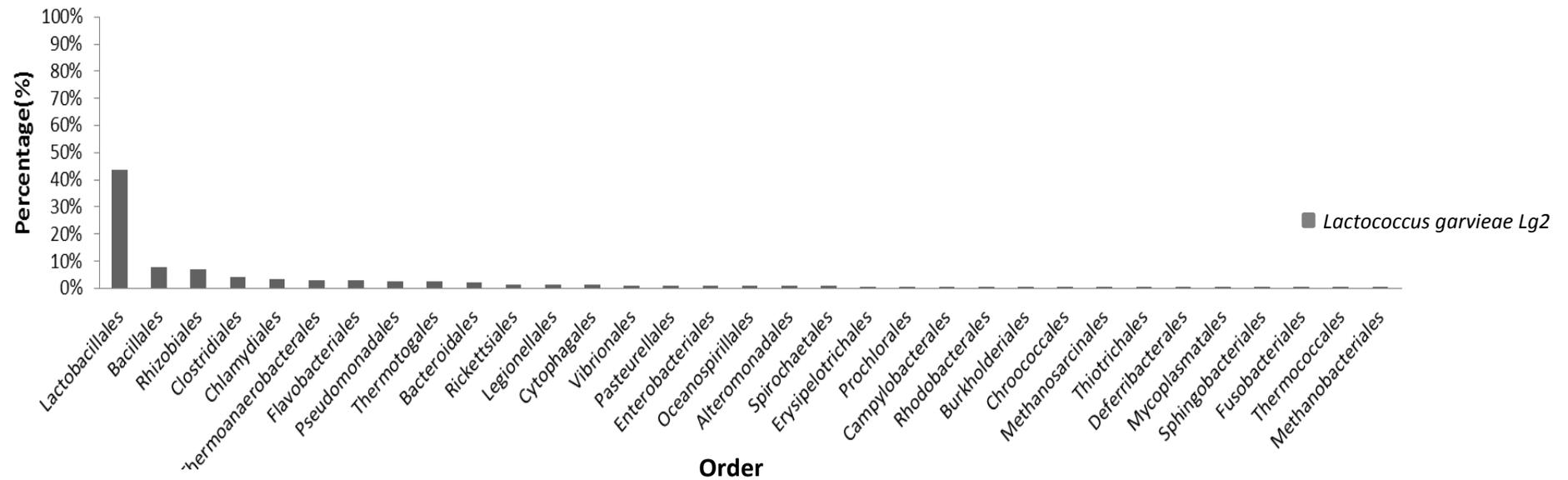
圖十九、*Lactococcus garvieae* Lg2 在門層級的分類結果

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在門層級的分類上各別被分至那些門中，縱軸顯示這些被分類到的門各別佔了多少的百分比。百分比的算法是利用被分類到該門的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



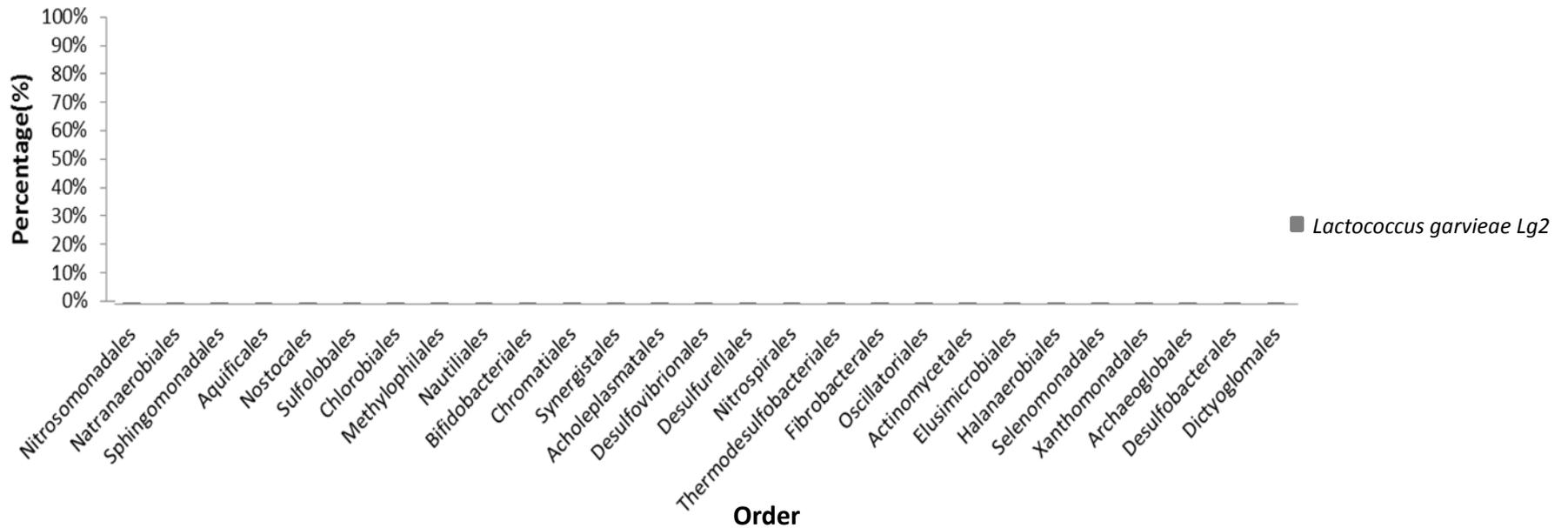
圖二十、*Lactococcus garvieae* Lg2 在綱層級的分類結果

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在綱層級的分類上各別被分至那些綱中，縱軸顯示這些被分類到的綱各別佔了多少的百分比。百分比的算法是利用被分類到該綱的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



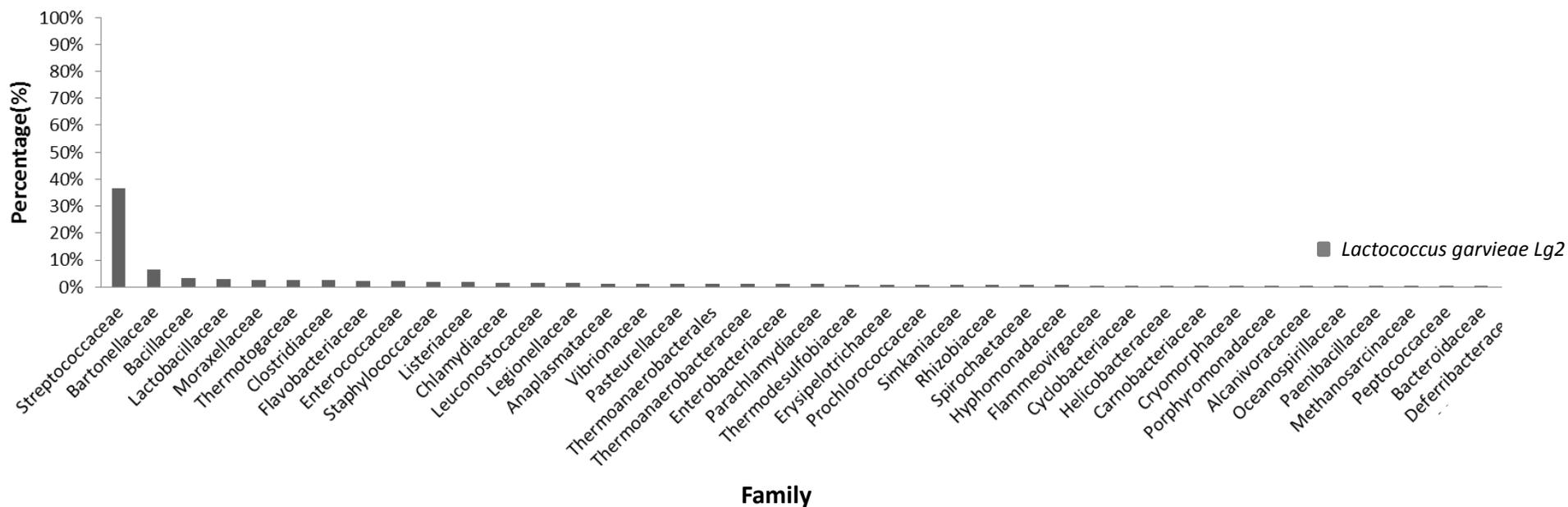
圖二十一、*Lactococcus garvieae* Lg2 在目層級的分類結果

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在目層級的分類上各別被分至那些目中，縱軸顯示這些被分類到的各別佔了多少的百分比。百分比的算法是利用被分類到該目的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



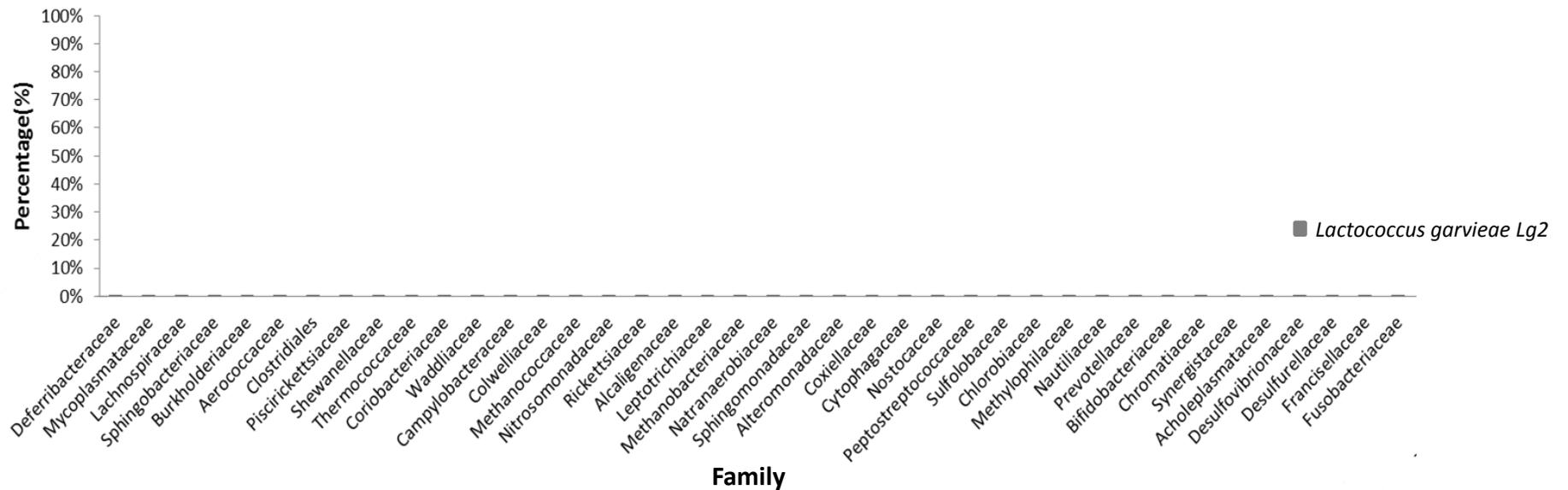
圖二十一、*Lactococcus garvieae* Lg2 在目層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在目層級的分類上各別被分至那些目中，縱軸顯示這些被分類到的各別佔了多少的百分比。百分比的算法是利用被分類到該目的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



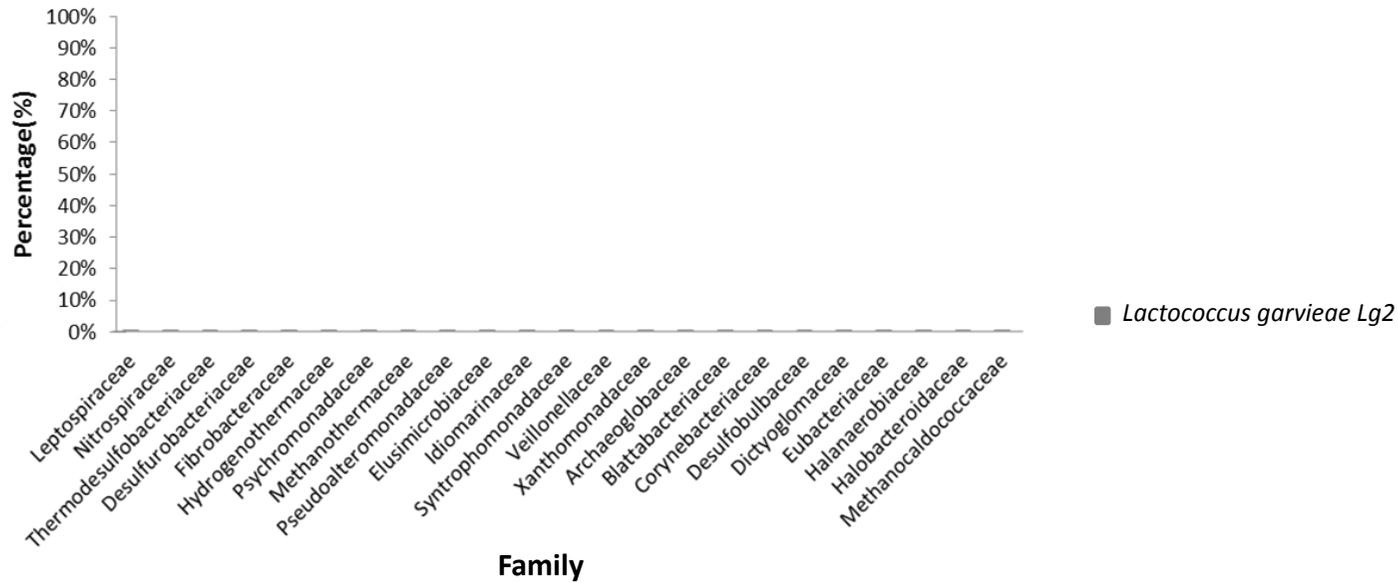
圖二十二、*Lactococcus garvieae* Lg2 在科層級的分類結果

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



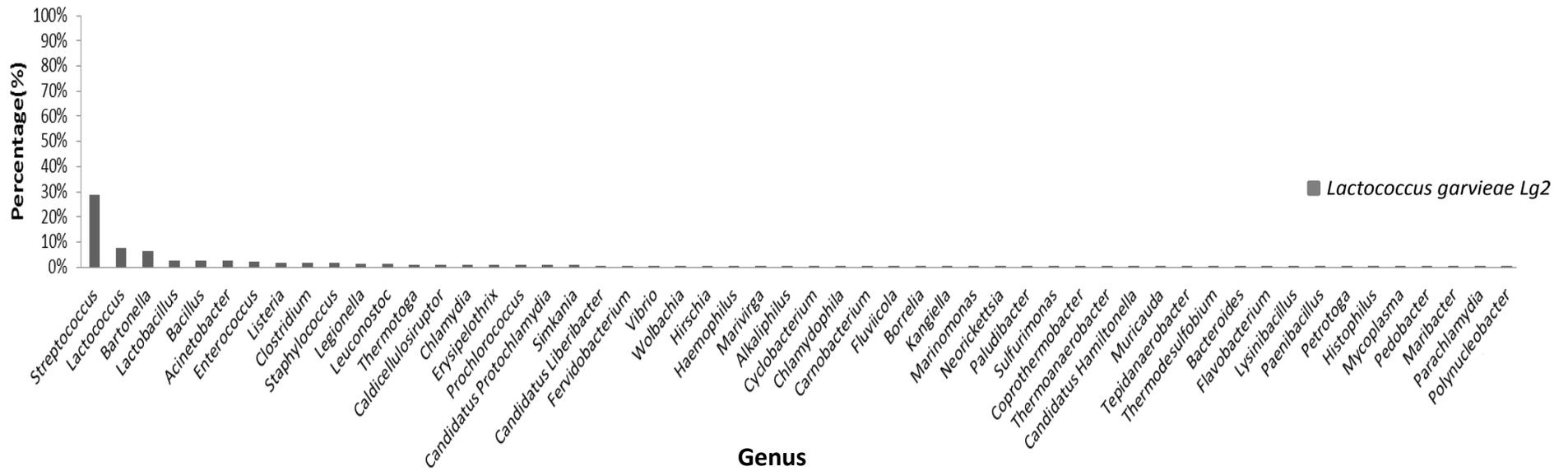
圖二十二、*Lactococcus garvieae* Lg2 在科層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



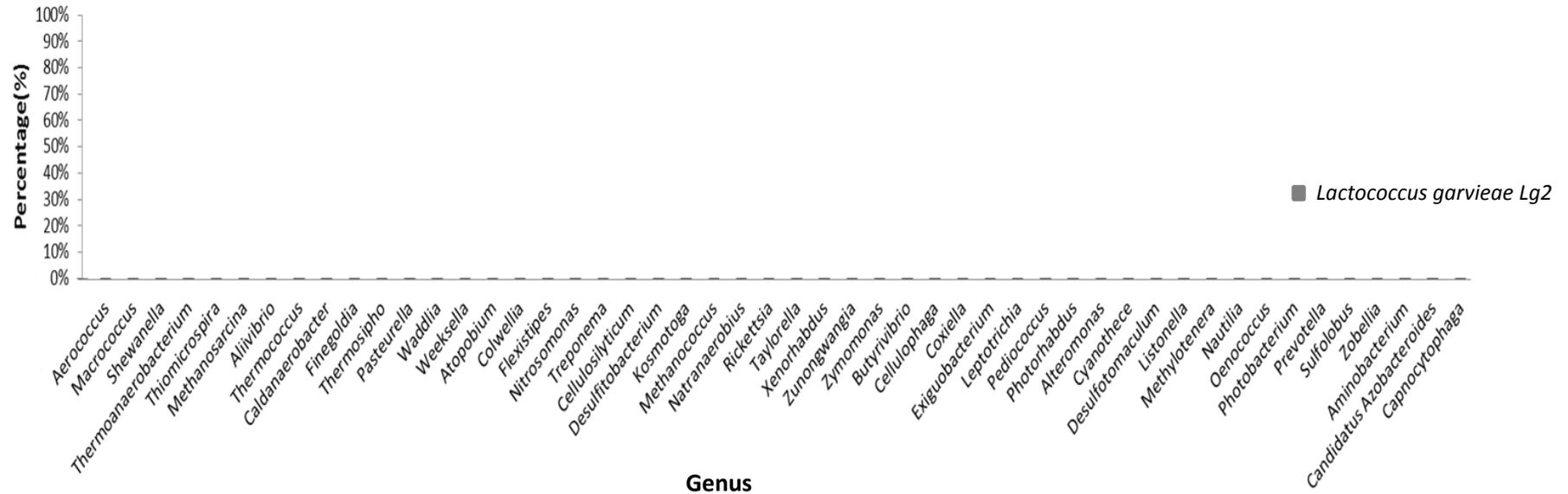
圖二十二、*Lactococcus garvieae* Lg2 在科層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在科層級的分類上各別被分至那些科中，縱軸顯示這些被分類到的科各別佔了多少的百分比。百分比的算法是利用被分類到該科的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



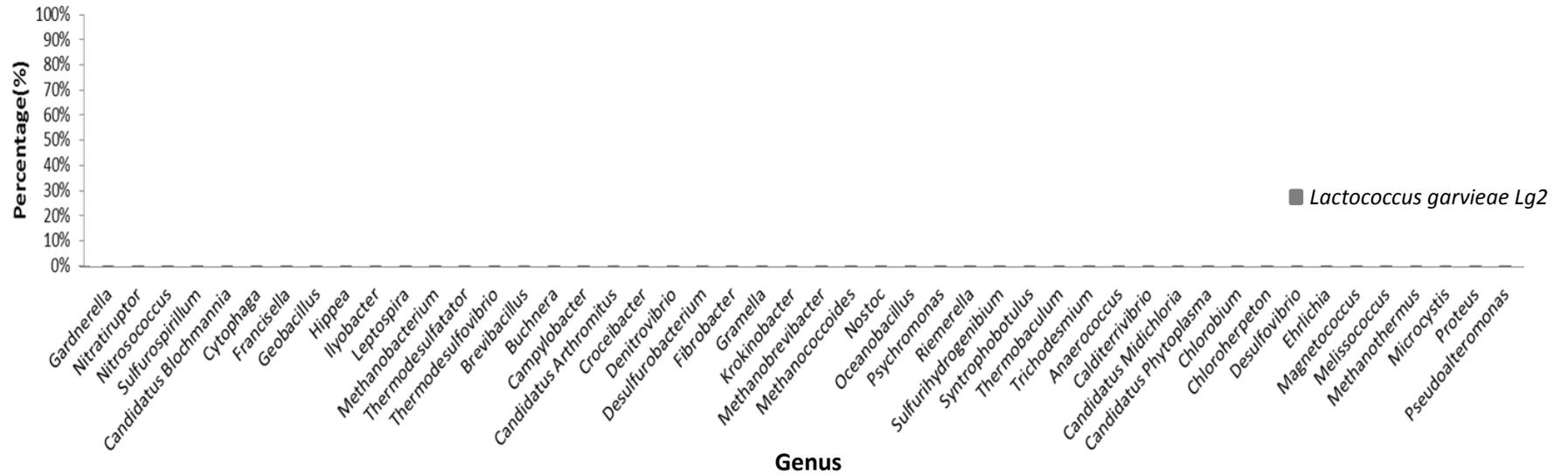
圖二十三、*Lactococcus garvieae* Lg2 在屬層級的分類結果

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



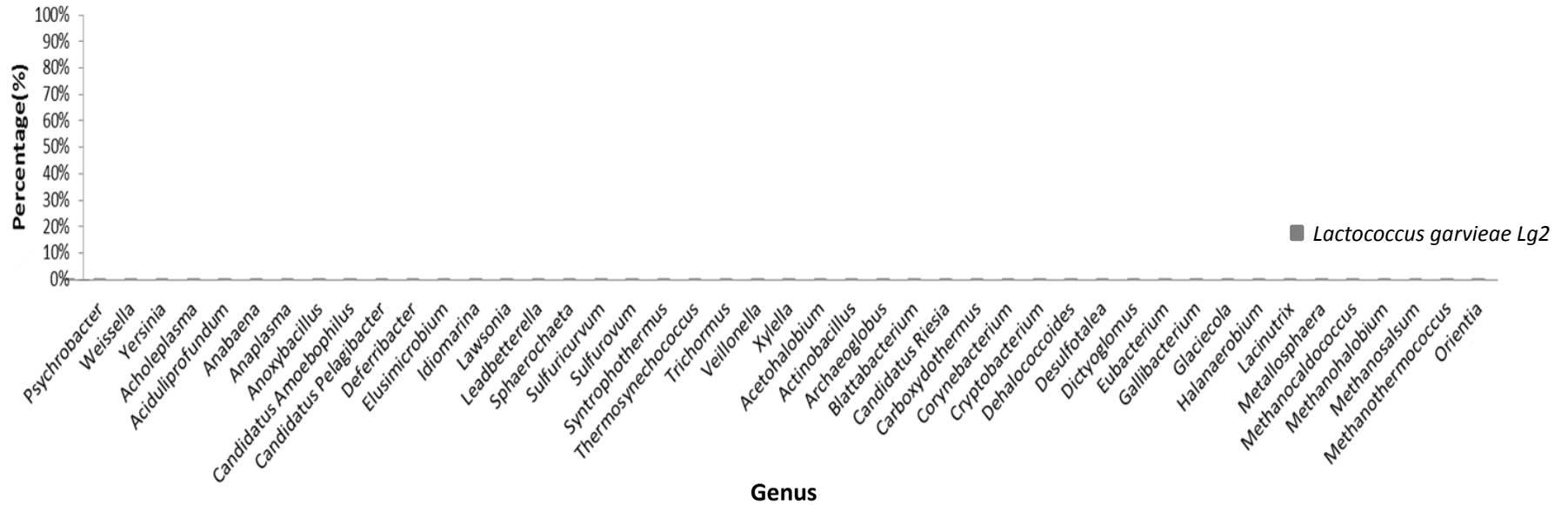
圖二十三、*Lactococcus garvieae* Lg2 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



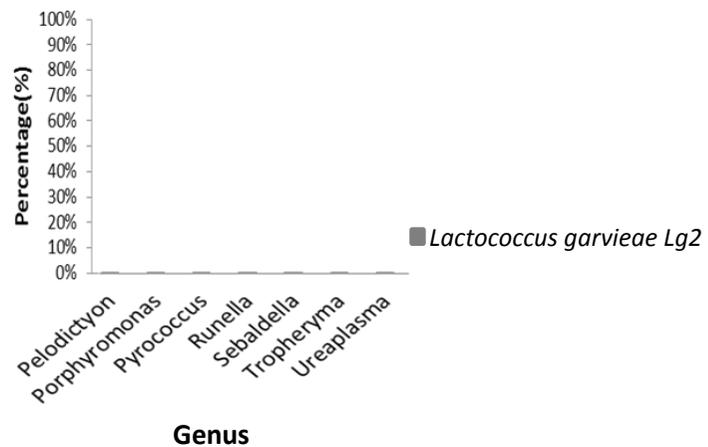
圖二十三、*Lactococcus garvieae* Lg2 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



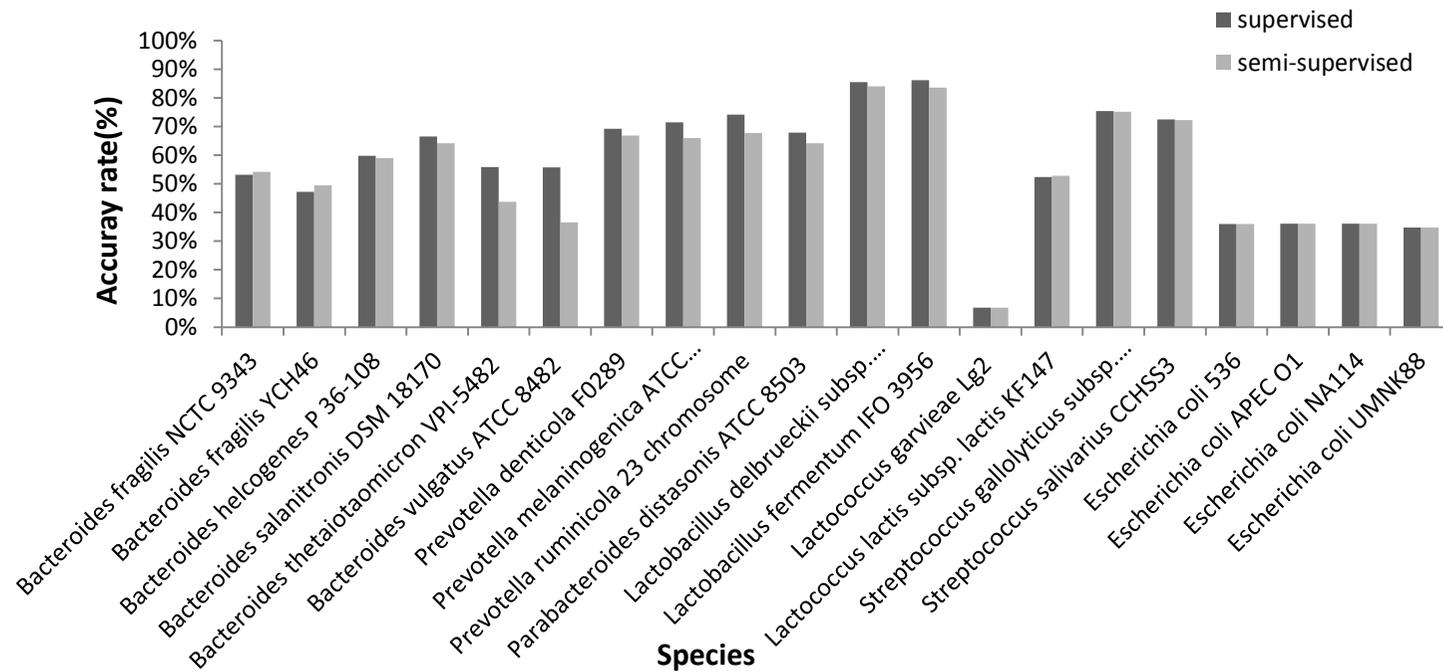
圖二十三、*Lactococcus garvieae* Lg2 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



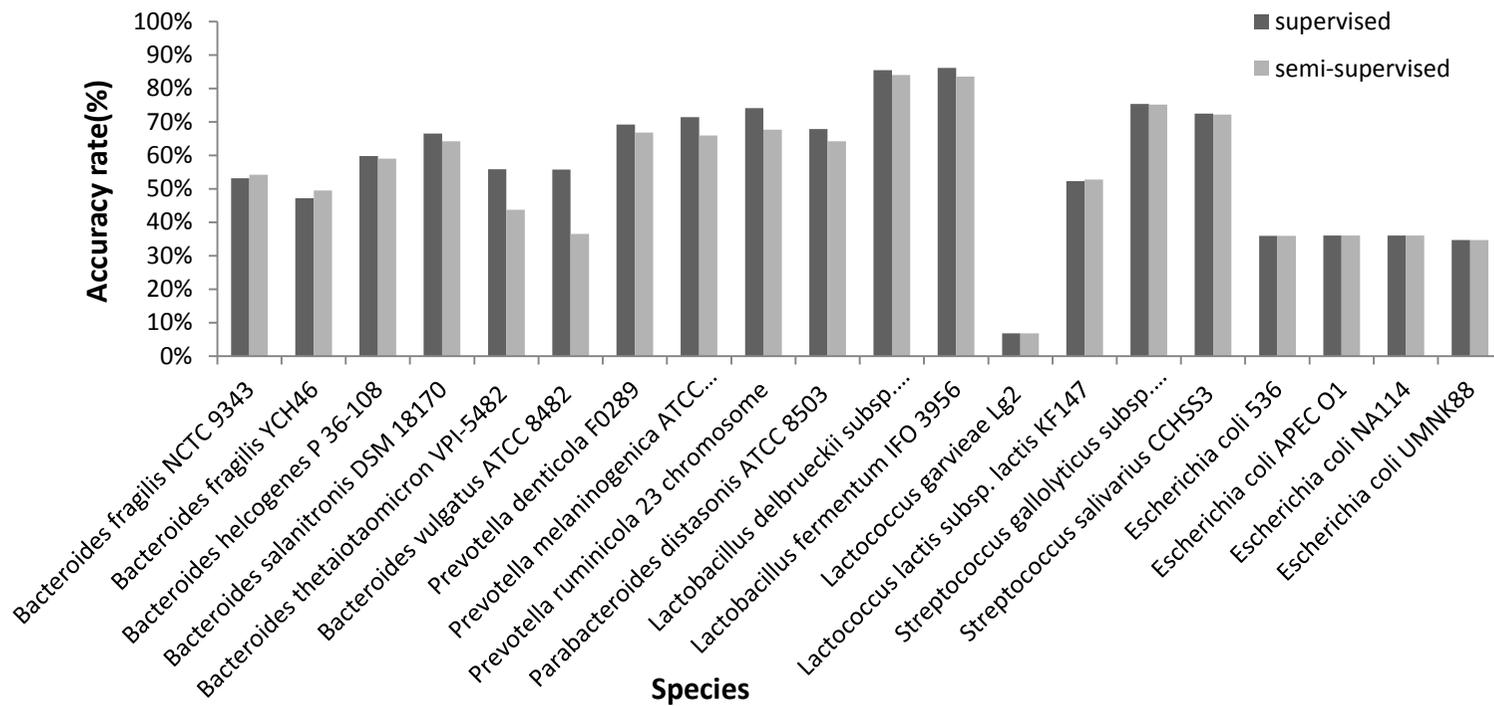
圖二十三、*Lactococcus garvieae* Lg2 在屬層級的分類結果(續)

圖的橫軸顯示測試的五千條 *Lactococcus garvieae* Lg2 序列在屬層級的分類上各別被分至那些屬中，縱軸顯示這些被分類到的屬各別佔了多少的百分比。百分比的算法是利用被分類到該屬的 *Lactococcus garvieae* Lg2 序列數目除以 *Lactococcus garvieae* Lg2 測試序列的總數。



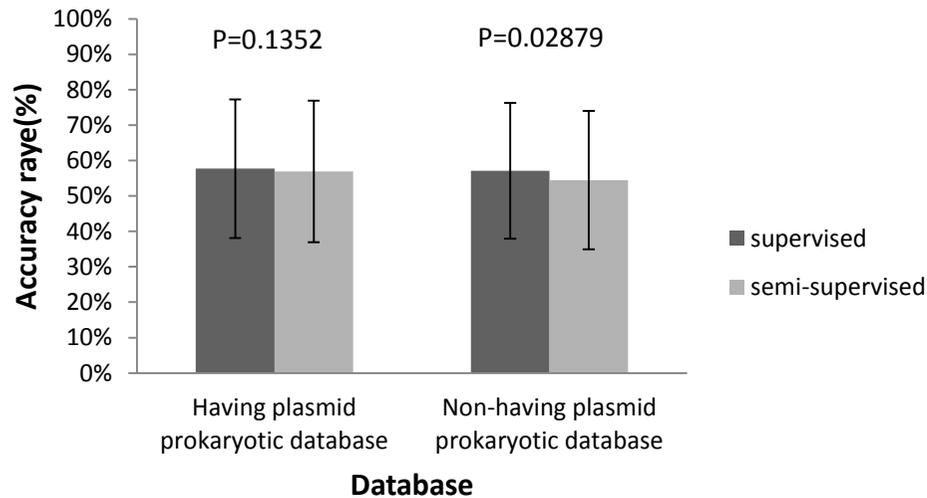
圖二十四、未加質體的原核生物序列特徵資料庫測試

利用 supervised 和 semi-supervised 兩種不同的操作方法，測試未加入質體序列的序列特徵資料庫分類二十種不同 strain 的腸胃道細菌序列的結果。圖的橫軸為二十種不同 strain 的腸胃道細菌序列名稱，縱軸的分類準確度為各原核生物序列被準確分類到屬的數目除以該原核生生物測試序列的總數。



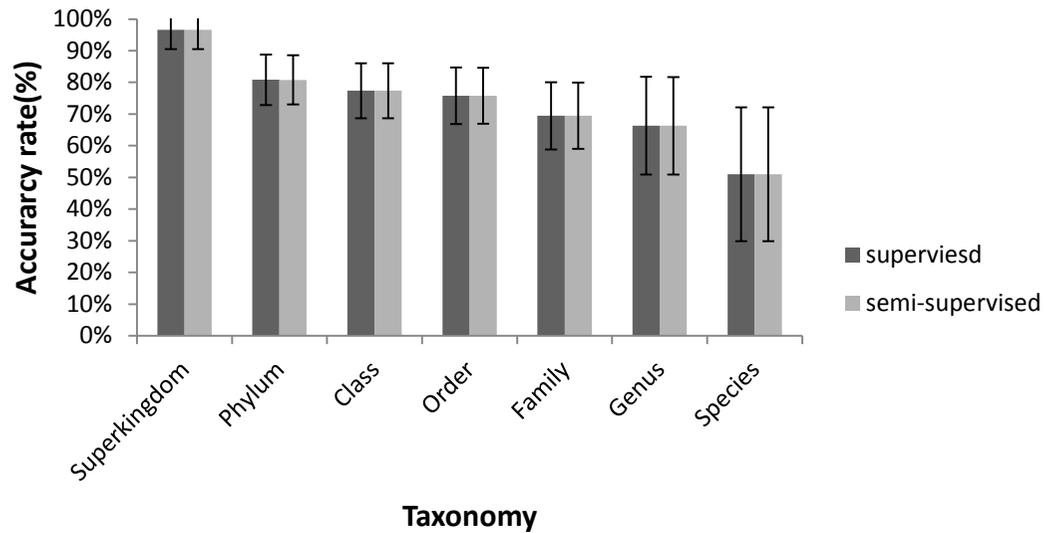
圖二十五、加入質體的原核生物序列特徵資料庫測試

利用 supervised 和 semi-supervised 兩種不同的操作方法，測試加入質體序列序列特徵資料庫分類二十種不同 strain 的腸胃道屬細菌序列的結果。圖的橫軸為二十種不同 strain 的腸胃道屬原核生物序列名稱，縱軸的分類準確度為各原核生物序列被準確分類到屬的數目除以該原核生生物測試序列的總數。



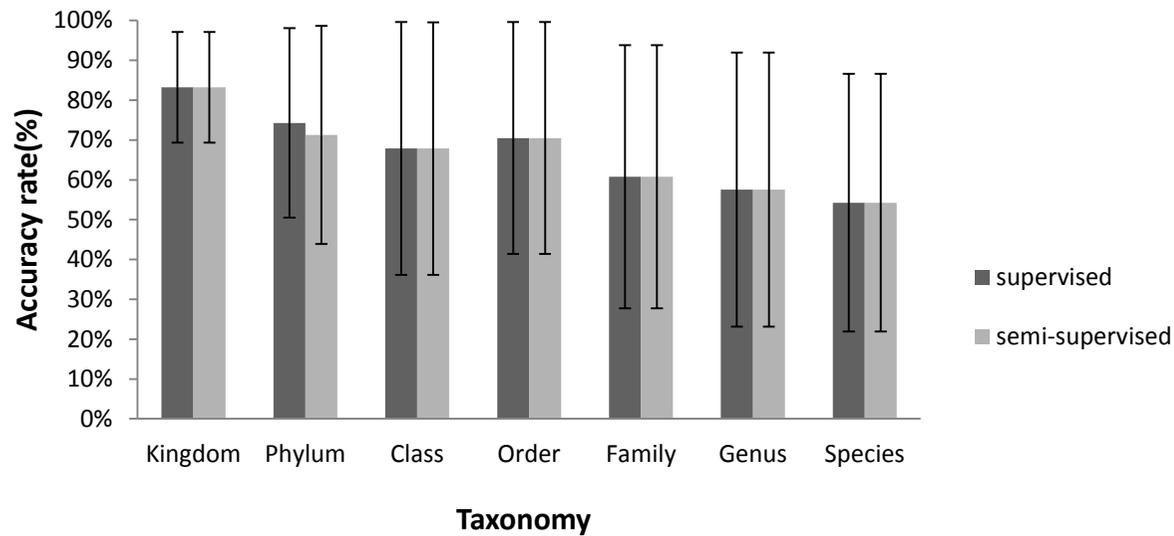
圖二十六、原核生物資料庫的平均分類準確度

圖的橫軸為有無質體序列的原核生物序列特徵資料庫，縱軸為這兩種資料庫分類二十種不同 strain 腸胃道細菌序列的平均準確度。P 值為 t-test 的比較結果，比較在 supervised 和 semi-supervised 兩種不同的操作方法下，腸胃道細菌序列的分類準確度於屬的層級上是否有所差異，結果顯示有質體序列的資料庫兩種測試方法上無顯著差異，P 值大於 0.05；而有質體序列的資料庫兩種測試方法上有顯著差異，P 值小於 0.05。



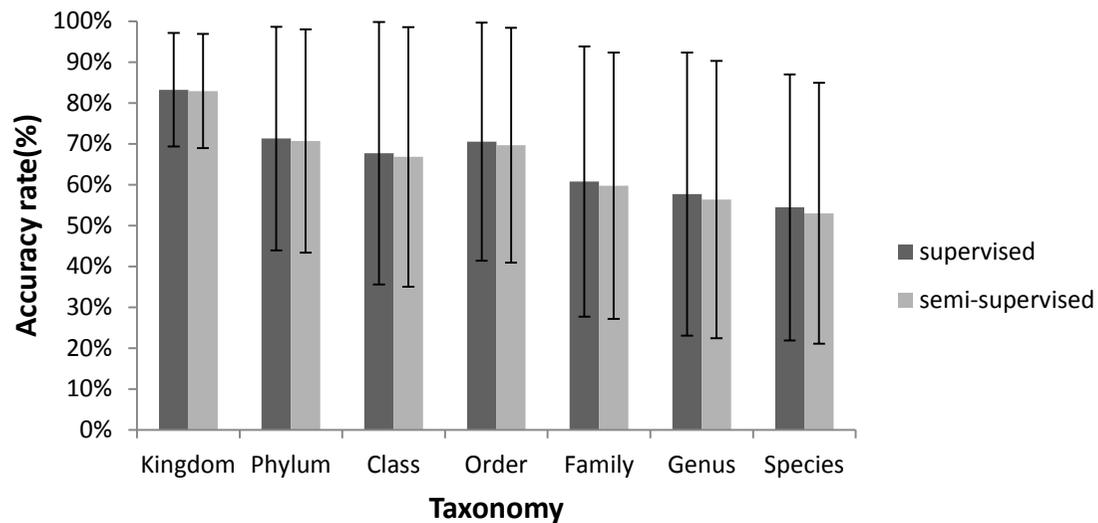
圖二十七、三十種原核生物序列測試

利用 supervised 和 semi-supervised 兩種不同的操作方法，測試無質體序列的原核生物序列特徵資料庫分類三十個不同種的腸胃道細菌序列的準確度，計算這些序列在各層級的平均準確度。圖的橫軸為分類層級，縱軸為各階層分類的平均準確度。



圖二十八、真核生物序列特徵資料庫測試

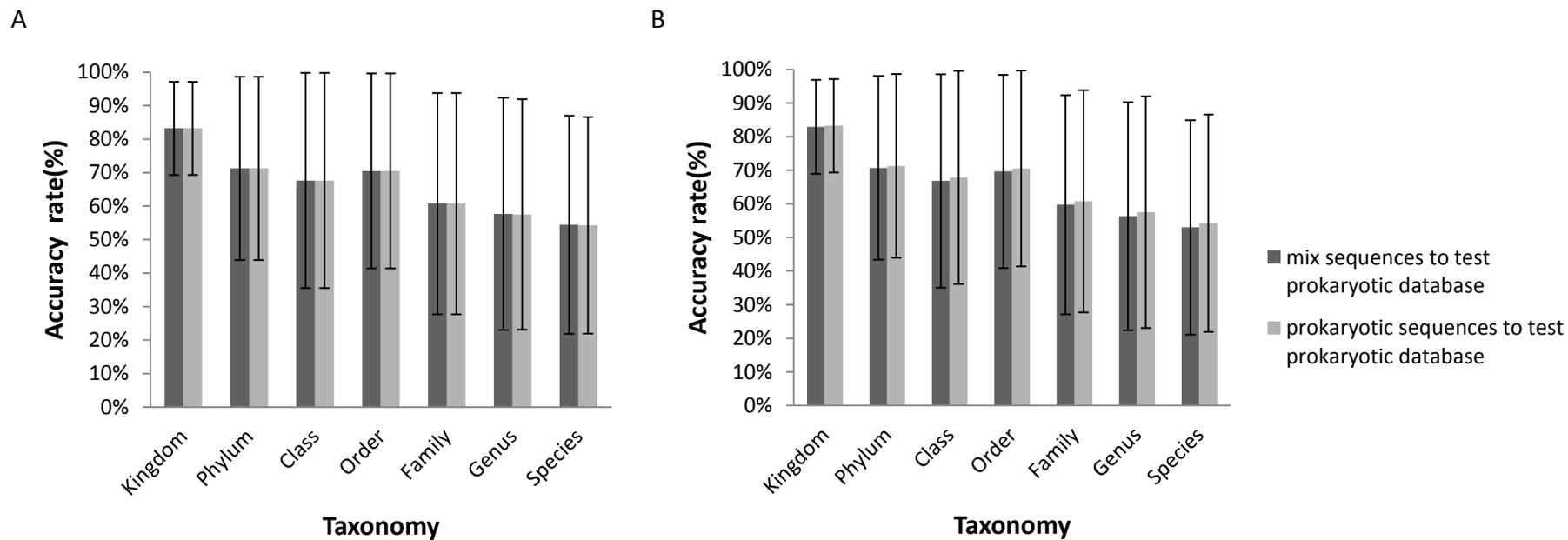
利用 supervised 和 semi-supervised 兩種不同的操作方法，測試真核生物序列特徵資料庫分類三十個不同種類的腸胃道細菌序列的準確度。圖的橫軸為分類層級，縱軸為各階層分類的平均準確度。



圖二十九、混合序列測試真核生物序列特徵資料庫

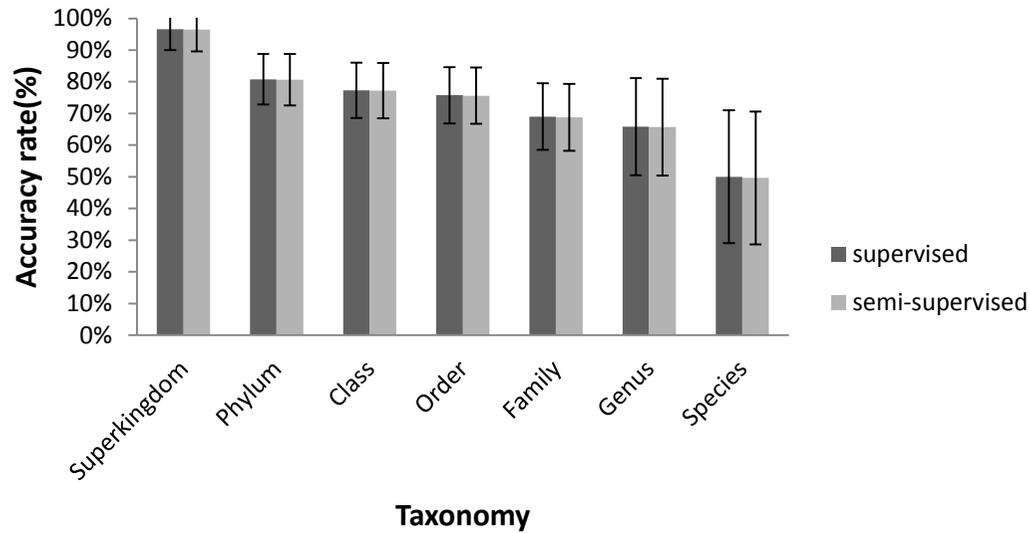
利用 supervised 和 semi-supervised 兩種不同的操作方法，測試真核生物序列特徵資料庫分類混合序列的準確度。圖

的橫軸為分類層級，縱軸為各階層分類的平均準確度。



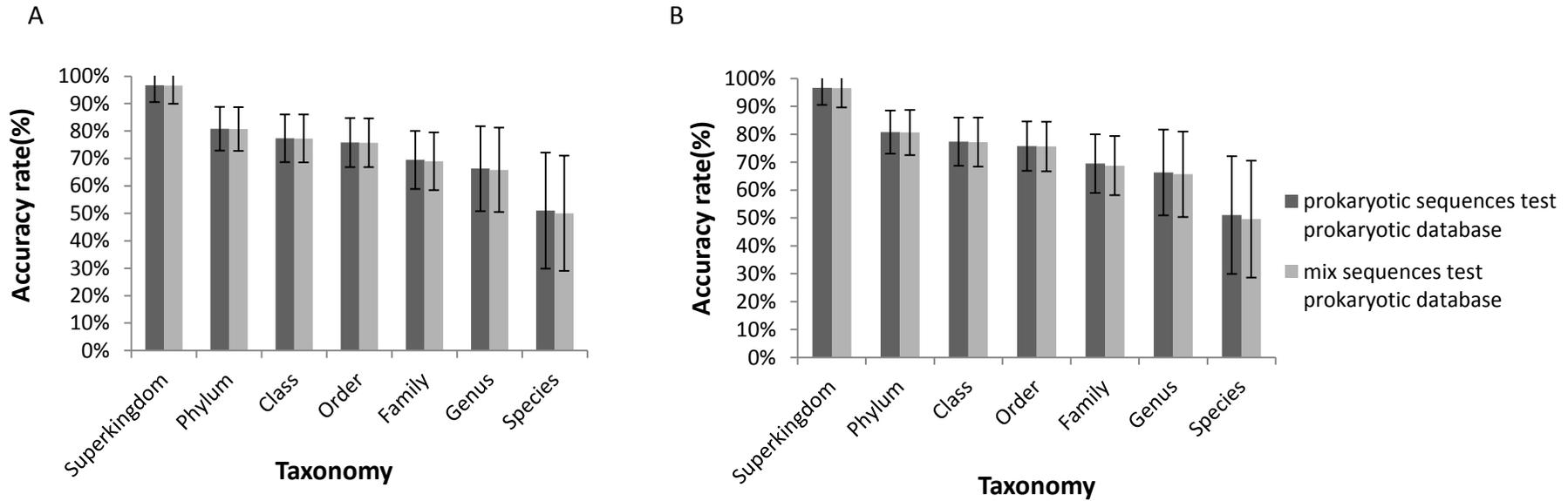
圖三十、比較混合與真核生物序列分類結果

比較在 supervised(A)以及在 semi-supervised(B)的狀態下，真核生物序列特徵資料庫分類混合序列與真核序列分類準確度的差異。圖的橫軸為分類層級，縱軸為各階層分類的平均準確度。



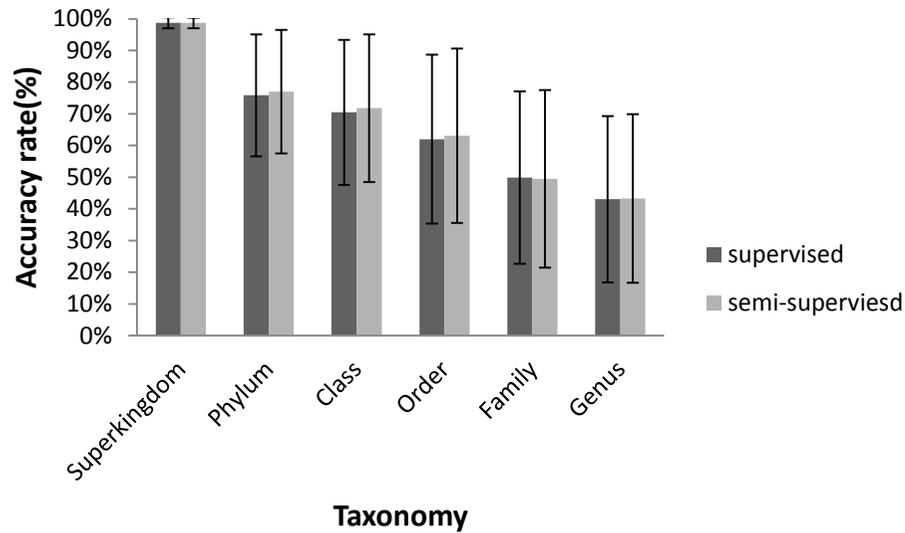
圖三十一、混合序列測試未加質體的原核生物序列特徵資料庫

利用 supervised 和 semi-supervised 兩種不同的操作方法，測試無質體的原核生物序列特徵資料庫分類混合序列的準確度。圖的橫軸為分類層級，縱軸為各階層分類的平均準確度。



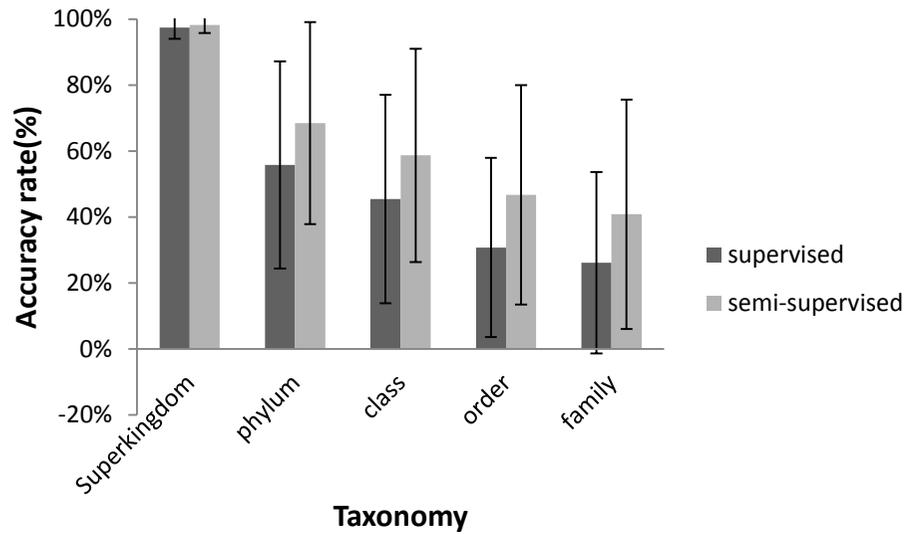
圖三十二、比較混合與原核生物序列分類結果

比較在 supervised(A)以及在 semi-supervised(B)的狀態下，原核生物序列特徵資料庫分類混合序列與原核序列分類準確度的差異。圖的橫軸為分類層級，縱軸為各階層分類的平均準確度。



圖三十三、三十個種層級不存於資料庫的序列類測試

利用 supervised 和 semi-supervised 兩種不同的操作方法，測試無質體序列的原核生物序列特徵資料庫分類三十種資料庫不存在的原核序列測試結果。圖的橫軸為分類層級，縱軸為各階層分類的平均準確度。



圖三十四、三十個屬層級不存於資料庫的序列測試

利用 supervised 和 semi-supervised 兩種不同的操作方法，測試無質體序列的原核生物序列特徵資料庫分類三十個屬層級不存於資料的原核序列測試結果。圖的橫軸為分類層級，縱軸為各階層分類的平均準確度。

表一、原核生物序列特徵資料庫所包含的分類層級

Kingdom					
Bacteria					
Archaea					
Phylum					
Acidobacteria	Chrysiogenetes	Euryarchaeota	Nitrospirae	Thermodesulfobacteria	
Actinobacteria	Crenarchaeota	Fibrobacteres	Planctomycetes	Thermotogae	
Aquificae	Cyanobacteria	Firmicutes	Proteobacteria	Verrucomicrobia	
Bacteroidetes	Deferribacteres	Fusobacteria	Spirochaetes		
Chlamydiae	Deinococcus-Thermus	Gemmatimonadetes	Synergistetes		
Chlorobi	Dictyoglomi	Korarchaeota	Tenericutes		
Chloroflexi	Elusimicrobia	Nanoarchaeota	Thaumarchaeota		
Class					
Acidobacteriia	Chlamydiia	Deltaproteobacteria	Gemmatimonadetes	Nitrospira	Thermomicrobia
Actinobacteria	Chlorobia	Dictyoglomia	Gloeobacteria	Opitutae	Thermoplasmata
Alphaproteobacteria	Chloroflexi	Elusimicrobia	Halobacteria	Planctomycetia	Thermoprotei
Anaerolineae	Chrysiogenetes	Epsilonproteobacteria	Methanobacteria	Solibacteres	Thermotogae
Aquificae	Clostridia	Erysipelotrichi	Methanococci	Sphingobacteriia	Verrucomicrobiae
Archaeoglobi	Cytophagia	Fibrobacteria	Methanomicrobia	Spirochaetia	
Bacilli	Deferribacteres	Flavobacteriia	Methanopyri	Synergistia	
Bacteroidia	Dehalococcoidetes	Fusobacteriia	Mollicutes	Thermococci	
Betaproteobacteria	Deinococci	Gammaproteobacteria	Negativicutes	Thermodesulfobacteria	
Order					
Acholeplasmatales	Cenarchaeales	Enterobacteriales	Methanosarcinales	Puniceicoccales	Thermodesulfobacteriales
Acidilobales	Chlamydiales	Entomoplasmatales	Methylacidiphilales	Rhizobiales	Thermomicrobiales
Acidimicrobiales	Chlorobiales	Erysipelotrichales	Methylococcales	Rhodobacterales	Thermoplasmatales
Acidithiobacillales	Chloroflexales	Fibrobacterales	Methylophilales	Rhodocyclales	Thermoproteales
Acidobacteriales	Chromatiales	Flavobacteriales	Mycoplasmatales	Rhodospirillales	Thermotogales
Actinomycetales	Chroococcales	Fusobacteriales	Myxococcales	Rickettsiales	Thiotrichales
Aeromonadales	Chrysiogenales	Gallionellales	Natranaerobiales	Rubrobacterales	Verrucomicrobiales
Alteromonadales	Clostridiales	Gemmatimonadales	Nautiliales	Selenomonadales	Vibrionales
Anaerolineales	Coriobacteriales	Gloeobacterales	Neisseriales	Solibacterales	Xanthomonadales
Aquificales	Cytophagales	Halanaerobiales	Nitrosomonadales	Solirubrobacterales	
Archaeoglobales	Deferribacteriales	Halobacteriales	Nitrosopumilales	Sphaerobacterales	
Bacillales	Deinococcales	Herpetosiphonales	Nitrospirales	Sphingobacteriales	
Bacteroidales	Desulfarculales	Hydrogenophilales	Nostocales	Sphingomonadales	
Bacteroidetes Order II.	Desulfobacteriales	Lactobacillales	Oceanospirillales	Spirochaetales	
Incertae sedis					
Bdellovibrionales	Desulfovibrionales	Legionellales	Oscillatoriales	Sulfolobales	
Bifidobacteriales	Desulfurellales	Methanobacteriales	Parvularculales	Synergistales	
Burkholderiales	Desulfurococcales	Methanocellales	Pasteurellales	Syntrophobacteriales	
Campylobacteriales	Desulfuromonadales	Methanococcales	Planctomycetales	Themales	
Cardiobacteriales	Dictyoglomales	Methanomicrobiales	Prochlorales	Thermoanaerobacterales	
Caulobacterales	Elusimicrobiales	Methanopyrales	Pseudomonadales	Thermococcales	

表一、原核生物序列特徵資料庫所包含的分類層級(續)

Family					
Acetobacteraceae	Chloroflexaceae	Fibrobacteraceae	Methanococcaceae	Phyllobacteriaceae	Syntrophaceae
Acholeplasmataceae	Chromatiaceae	Flammeovirgaceae	Methanocorpusculaceae	Picrophilaceae	Syntrophobacteraceae
Acidaminococcaceae	Chrysiogenaceae	Flavobacteriaceae	Methanomicrobiaceae	Piscirickettsiaceae	Syntrophomonadaceae
Acidilobaceae	Clostridiaceae	Francisellaceae	Methanopyraceae	Planctomycetaceae	Thermaceae
Acidimicrobiaceae	Clostridiales Family XI. Incertae Sedis	Frankiaceae	Methanosaetaceae	Polyangiaceae	Thermoanaerobacteraceae
Acidithiobacillaceae	Clostridiales Family XVII. Incertae Sedis	Fusobacteriaceae	Methanosarcinaceae	Porphyromonadaceae	Thermoanaerobacterales Family III. Incertae Sedis
Acidobacteriaceae	Clostridiales Family XVIII. Incertae Sedis	Gallionellaceae	Methanospirillaceae	Prevotellaceae	Thermoanaerobacterales Family IV. Incertae Sedis
Acidothermaceae	Colwelliaceae	Gemmatimonadaceae	Methanothermaceae	Prochlorococcaceae	Thermococcaceae
Actinomycetaceae	Comamonadaceae	Geobacteraceae	Methylacidiphilaceae	Promicromonosporaceae	Thermodesulfobacteriaceae
Aerococcaceae	Conexibacteraceae	Geodermatophilaceae	Methylobacteriaceae	Propionibacteriaceae	Thermodesulfobiaceae
Aeromonadaceae	Coriobacteriaceae	Glycomycetaceae	Methylococcaceae	Pseudoalteromonadaceae	Thermofilaceae
Alcaligenaceae	Corynebacteriaceae	Gordoniaceae	Methylophilaceae	Pseudomonadaceae	Thermomicrobiaceae
Alcanivoracaceae	Coxiellaceae	Hahellaceae	Microbacteriaceae	Pseudonocardiaceae	Thermomonosporaceae
Alicyclobacillaceae	Cryomorphaceae	Halanaerobiaceae	Micrococcaceae	Psychromonadaceae	Thermoplasmataceae
Alteromonadaceae	Cyclobacteriaceae	Halobacteriaceae	Micromonosporaceae	Puniceicoccaceae	Thermoproteaceae
Anaerolineaceae	Cystobacteraceae	Halobacteroidaceae	Moraxellaceae	Pyrodictiaceae	Thermotogaceae
Anaplasmataceae	Cytophagaceae	Halomonadaceae	Mycobacteriaceae	Rhizobiaceae	Trueperaceae
Aquificaceae	Deferribacteraceae	Halothiobacillaceae	Mycoplasmataceae	Rhodobacteraceae	Tsakamurellaceae
Archaeoglobaceae	Deinococcaceae	Helicobacteraceae	Myxococcaceae	Rhodocyclaceae	Veillonellaceae
Bacillaceae	Dermabacteraceae	Hellobacteriaceae	Nakamurellaceae	Rhodospirillaceae	Verrucomicrobiaceae
Bacteroidaceae	Dermacoccaceae	Herpetosiphonaceae	Natranaerobiaceae	Rhodothermaceae	Vibrionaceae
Bartonellaceae	Desulfarculaceae	Hydrogenophilaceae	Nautiliaceae	Rickettsiaceae	Waddliaceae
Bdellovibrionaceae	Desulfobacteraceae	Hydrogenothermaceae	Neisseriaceae	Rubrobacteraceae	Xanthobacteraceae
Beijerinckiaceae	Desulfobulbaceae	Hyphomicrobiaceae	Nitrosomonadaceae	Ruminococcaceae	Xanthomonadaceae
Beutenbergiaceae	Desulfohalobiaceae	Hyphomonadaceae	Nitrosopumilaceae	Sanguibacteraceae	
Bifidobacteriaceae	Desulfomicrobiaceae	Idiomarinaceae	Nitrospiraceae	Saprospiraceae	
Blattabacteriaceae	Desulfovibrionaceae	Intrasporangiaceae	Nocardiaceae	Segniliparaceae	
Brachyspiraceae	Desulfurellaceae	Jonesiaceae	Nocardiodaceae	Shewanellaceae	
Bradyrhizobiaceae	Desulfurobacteriaceae	Kineosporiaceae	Nocardiopsaceae	Simkaniaceae	
Brucellaceae	Desulfurococcaceae	Koferiaceae	Nostocaceae	Solibacteraceae	
Burkholderiaceae	Dictyoglomaceae	Lachnospiraceae	Oceanospirillaceae	Sphaerobacteraceae	
Campylobacteraceae	Ectothiorhodospiraceae	Lactobacillaceae	Opitutaceae	Sphingobacteriaceae	
Cardiobacteriaceae	Elusimicrobiaceae	Legionellaceae	Oxalobacteraceae	Sphingomonadaceae	
Carnobacteriaceae	Enterobacteriaceae	Leptospiraceae	Paenibacillaceae	Spirochaetaceae	
Catenulisporaceae	Enterococcaceae	Leptotrichiaceae	Parachlamydiaceae	Staphylococcaceae	
Caulobacteraceae	Entomoplasmataceae	Leuconostocaceae	Parvularculaceae	Streptococcaceae	
Cellulomonadaceae	Erysipelotrichaceae	Listeriaceae	Pasteurellaceae	Streptomycetaceae	
Cenarchaeaceae	Erythrobacteraceae	Methanobacteriaceae	Pelobacteraceae	Streptosporangiaceae	
Chlamydiaceae	Eubacteriaceae	Methanocaldococcaceae	Peptococcaceae	Sulfobacillaceae	
Chlorobiaceae	Ferrimonadaceae	Methanocellaceae	Peptostreptococcaceae	Synergistaceae	

表一、原核生物序列特徵資料庫所包含的分類層級(續)

Genus					
Acaryochloris	Beijerinckia	Chloroherpeton	Ethanoligenens	Hyphomonas	Methanohalophilus
Acetobacter	Beutenbergia	Chromobacterium	Eubacterium	Idiomarina	Methanoplanus
Acetohalobium	Bifidobacterium	Chromohalobacter	Exiguobacterium	Ignicoccus	Methanopyrus
Acholeplasma	Blattabacterium	Citrobacter	Ferrimonas	Ignisphaera	Methanoregula
Achromobacter	Bordetella	Clavibacter	Ferroglobus	Ilyobacter	Methanosaeta
Acidaminococcus	Borrelia	Clostridium	Fervidobacterium	Intrasporangium	Methanosalsum
Acidianus	Brachybacterium	Collimonas	Fibrobacter	Isoptricola	Methanosarcina
Acidilobus	Brachyspira	Colwellia	Finegoldia	Isosphaera	Methanosphaera
Acidimicrobium	Bradyrhizobium	Comamonas	Flavobacterium	Jannaschia	Methanosphaerula
Acidiphilium	Brevibacillus	Conexibacter	Flexistipes	Janthinobacterium	Methanospirillum
Acidithiobacillus	Brevundimonas	Coprothermobacter	Fluviicola	Jonesia	Methanothermobacter
Acidobacterium	Brucella	Coraliomargarita	Francisella	Kangiella	Methanothermococcus
Acidothermus	Buchnera	Coriobacterium	Frankia	Ketogulonicigenium	Methanothermus
Acidovorax	Burkholderia	Corynebacterium	Fusobacterium	Kineococcus	Methanotorris
Aciduliprofundum	Butyrivibrio	Coxiella	Coxiella	Gallibacterium	Methylacidiphilum
Acinetobacter	Caldanaerobacter	Croceibacter	Gallionella	Kocuria	Methylibium
Actinobacillus	Caldicellulosiruptor	Cronobacter	Gardnerella	Kosmotoga	Methylbacillus
Actinosynnema	Calditerrivibrio	Cryptobacterium	Gemmatimonas	Kribbella	Methylobacterium
Aerococcus	Caldivirga	Cupriavidus	Geobacillus	Geobacter	Methylocella
Aeromonas	Campylobacter	Cyanothece	Geobacter	Geodermatophilus	Methylococcus
Aeropyrum	Candidatus Accumulibacter	Cyclobacterium	Glaciicola	Gloeobacter	Methylococcus
Aggregatibacter	Candidatus Amoebophilus	Cytophaga	Gloebacter	Gluconacetobacter	Methylotenera
Agrobacterium	Candidatus Arthromitus	Dechloromonas	Gluconobacter	Gluconobacter	Methylotenus
Akkermansia	Candidatus Azobacteroides	Deferribacter	Gordonia	Laribacter	Microbacterium
Albidiferax	Candidatus Blochmannia	Dehalococcoides	Gramella	Lawsonia	Micrococcus
Alcanivorax	Candidatus Carsonella	Dehalogenimonas	Granulibacter	Leadbetterella	Microcystis
Alicyclophilus	Candidatus Desulforudis	Deinococcus	Haemophilus	Legionella	Microlunatus
Alicyclobacillus	Candidatus Hamiltonella	Delftia	Hahella	Leptospira	Micromonospora
Aliivibrio	Candidatus Hodgkinia	Denitrovibrio	Halalkalicoccus	Leptothrix	Mobiluncus
Alkalilimnicola	Candidatus Korarchaeum	Desulfarculus	Halanaerobium	Leptotrichia	Moorella
Alkaliphilus	Candidatus Koribacter	Desulfatibacillum	Halobacterium	Leuconostoc	Moraxella
Allochromatium	Candidatus Liberibacter	Desulfatobacterium	Haloferax	Listeria	Muricauda
Alteromonas	Candidatus Midichloria	Desulfobacca	Halogeometricum	Listonella	Mycobacterium
Aminobacterium	Candidatus Moranella	Desulfobacterium	Halomicrobium	Lysinibacillus	Mycoplasma
Ammonifex	Candidatus Pelagibacter	Desulfobulbus	Halomonas	Macrococcus	Myxococcus
Amycolatopsis	Candidatus Phytoplasma	Desulfococcus	Halopiger	Magnetococcus	Nakamurella
Amycolicoccus	Candidatus Protochlamydia	Desulfohalobium	Haloquadratum	Magnetospirillum	Nanoarchaeum
Anabaena	Candidatus Puniceispirillum	Desulfomicrobium	Halorhabdus	Mahella	Natranaerobius
Anaerococcus	Candidatus Riesia	Desulfotalea	Halorhodospira	Maribacter	Natrialba
Anaerolinea	Candidatus Solibacter	Desulfotomaculum	Halorubrum	Maricoccus	Natronomonas
Anaeromyxobacter	Candidatus Sulcia	Desulfovibrio	Haloterrigena	Marinibacter	Nautilla
Anaplasma	Candidatus Tremblaya	Desulfurivibrio	Halothermothrix	Marinomonas	Neisseria
Anoxybacillus	Candidatus Zinderia	Desulfurivibrio	Halothiobacillus	Marinospira	Neorickettsia
Aquifex	Capnocytophaga	Desulfurococcus	Helicobacter	Marinobacter	Nitratifactor
Arcanobacterium	Carboxydotherrmus	Desulfurococcus	Heliobacterium	Marinomonas	Nitratiruptor
Archaeoglobus	Carnobacterium	Dichelobacter	Herbaspirillum	Marivirga	Nitrobacter
Arcobacter	Catenulispora	Dickeya	Herminimonas	Meiothermus	Nitrosococcus
Aromatoleum	Caulobacter	Dictyoglomus	Herpetosiphon	Melissococcus	Nitrosomonas
Arthrobacter	Cellulomonas	Dinoroseobacter	Hippea	Mesoplasma	Nitrosopumilus
Asticcacaulis	Cellulophaga	Dyadobacter	Hirschia	Mesohizobium	Nitrosospora
Atopobium	Cellulosilyticum	Edwardsiella	Histophilus	Metallosphaera	Nitrospirilla
Azoarcus	Cellvibrio	Eggerthella	Hydrogenobacter	Methanobrevibacter	Nocardia
Azorhizobium	Cenarchaeum	Ehrlichia	Hydrogenobacter	Methanocaldococcus	Nocardioideis
Azospirillum	Chelativorans	Elusimicrobium	Hydrogenobaculum	Methanocella	Nocardioideis
Azotobacter	Chitinophaga	Enterobacter	Hypothermus	Methanococcus	Nostoc
Bacillus	Chlamydia	Enterococcus	Hypomicrobium	Methanococcus	Novosphingobium
Bacteroides	Chlamydophila	Erwinia		Methanococcus	Oceanithermus
Bartonella	Chlorobaculum	Erythrobacter		Methanococcus	Oceanobacillus
Basfia	Chlorobium	Escherichia		Methanococcus	Ochrobactrum
Bdellovibrio	Chloroflexus			Methanococcus	Odoribacter

表一、原核生物序列特徵資料庫所包含的分類層級(續)

Genus					
Oenococcus	Prosthecochloris	Saccharophagus	Streptomyces	Thermodesulfobium	Weeksella
Oligotropha	Proteus	Saccharopolyspora	Streptosporangium	Thermodesulfovibrio	Weissella
Olsenella	Pseudoalteromonas	Salinibacter	Sulfobacillus	Thermofilum	Wigglesworthia
Opitutus	Pseudomonas	Salinispora	Sulfolobus	Thermomicrobium	Wolbachia
Orientia	Pseudonocardia	Salmonella	Sulfuricurvum	Thermomonospora	Wolinella
Paenibacillus	Pseudoxanthomonas	Sanguibacter	Sulfurihydrogenibium	Thermoplasma	Xanthobacter
Paludibacter	Psychrobacter	Sebaldella	Sulfurimonas	Thermoproteus	Xanthomonas
Pantoea	Psychromonas	Segniliparus	Sulfurospirillum	Thermosediminibacter	Xenorhabdus
Parabacteroides	Pusillimonas	Selenomonas	Sulfurovum	Thermosipho	Xylanimonas
Parachlamydia	Pyrobaculum	Serratia	Symbiobacterium	Thermosphaera	Xylella
Paracoccus	Pyrococcus	Shewanella	Synechococcus	Thermosynechococcus	Yersinia
Parvibaculum	Rahnella	Shigella	Synechocystis	Thermotoga	Zobellia
Parvularcula	Ralstonia	Sideroxydans	Syntrophobacter	Thermovibrio	Zunongwangia
Pasteurella	Ramlibacter	Simkania	Syntrophobotulus	Thermus	Zymomonas
Pectobacterium	Renibacterium	Sinorhizobium	Syntrophomonas	Thioalkalimicrobium	
Pediococcus	Rhizobium	Slackia	Syntrophothermus	Thioalkalivibrio	
Pedobacter	Rhodobacter	Sodalis	Syntrophus	Thiobacillus	
Pelobacter	Rhodococcus	Sorangium	Taylorella	Thiomicrospira	
Pelodictyon	Rhodomicrobium	Sphaerobacter	Tepidanaerobacter	Thiomonas	
Pelotomaculum	Rhodopirellula	Sphaerochaeta	Teredinibacter	Tolumonas	
Persephonella	Rhodopseudomonas	Sphingobacterium	Terriglobus	Treponema	
Petrotoga	Rhodospirillum	Sphingobium	Thauera	Trichodesmium	
Phenylobacterium	Rhodotherrmus	Sphingomonas	Thermaerobacter	Trichormus	
Photobacterium	Rickettsia	Sphingopyxis	Thermanaerovibrio	Tropheryma	
Photorhabdus	Riemerella	Spirochaeta	Thermincola	Truepera	
Picrophilus	Robiginitalea	Spirosoma	Thermoanaerobacter	Tsukamurella	
Pirellula	Roseiflexus	Stackebrandtia	Thermoanaerobacterium	Ureaplasma	
Planctomyces	Roseobacter	Staphylococcus	Thermobaculum	Variovorax	
Polaromonas	Rothia	Staphylothermus	Thermobifida	Veillonella	
Polynucleobacter	Rubrobacter	Starkeya	Thermobispora	Verminephrobacter	
Porphyromonas	Ruegeria	Stenotrophomonas	Thermococcus	Verrucosipora	
Prevotella	Ruminococcus	Stigmatella	Thermocrinis	Vibrio	
Prochlorococcus	Runella	Streptobacillus	Thermodesulfatator	Vulcanisaeta	
Propionibacterium	Saccharomonospora	Streptococcus	Thermodesulfobacterium	Waddlia	

表一、原核生物序列特徵資料庫所包含的分類層級(續)

Species					
[Cellvibrion] gilvus	Anaerolinea thermophila	Bifidobacterium bifidum	Campylobacter curvus	Chloroflexus aggregans	Cupriavidus pinatubonensis
[Clostridium] difficile	Anaeromyxobacter dehalogenans	Bifidobacterium dentium	Campylobacter fetus	Chloroflexus aurantiacus	Cupriavidus taiwanensis
[Clostridium] sticklandii	Anaeromyxobacter sp.	Bifidobacterium longum	Campylobacter hominis	Chloroflexus sp.	Cyanobacterium UCYN-A
Acaryochloris marina	Anaplasma centrale	Blattabacterium sp.	Campylobacter jejuni	Chloroherpeton thalassium	Cyanotheca sp.
Acetobacter pasteurianus	Anaplasma marginale	Bordetella avium	Campylobacter lari	Chromobacterium violaceum	Cyclobacterium marinum
Acetohalobium arabaticum	Anaplasma phagocytophilum	Bordetella bronchiseptica	Candidatus Accumulibacter phosphatis	Chromohalobacter salexigenis	Cytophaga hutchinsonii
Acholeplasma laidlawii	Anoxybacillus flavithermus	Bordetella parapertussis	Candidatus Amoebophilus asiaticus	Citrobacter koseri	Dechloromonas aromatica
Achromobacter xylosoxidans	Aquifex aeolicus	Bordetella pertussis	Candidatus Arthromitus sp. SFB-mouse	Citrobacter rodentium	Deferribacter desulfuricans
Acidaminococcus fermentans	Arcanobacterium haemolyticum	Bordetella petrii	Candidatus Azobacteroides pseudotrichonymphae	Clavibacter michiganensis	Dehalococcoides ethenogenes
Acidianus hospitalis	Archaeoglobus fulgidus	Borrelia afzelii	Candidatus Blochmannia pennsylvanicus	Clostridiales genomsp. BVAB3	Dehalococcoides sp.
Acidilobus saccharovorans	Archaeoglobus profundus	Borrelia bavariensis	Candidatus Blochmannia vafer	Clostridium acetobutylicum	Dehalogenimonas lykanthropopellens
Acidimicrobium ferrooxidans	Archaeoglobus veneficus	Borrelia bissettii	Candidatus Carsonella ruddii	Clostridium beijerinckii	Deinococcus deserti
Acidiphilium cryptum	Arcobacter butzleri	Borrelia burgdorferi	Candidatus Desulfurella audaxviator	Clostridium botulinum	Deinococcus geothermalis
Acidiphilium multivorum	Arcobacter nitrofigillis	Borrelia duttonii	Candidatus Hamiltonella defensa	Clostridium cellulolyticum	Deinococcus maricopenis
Acidithiobacillus caldus	Aromatoleum aromaticum	Borrelia hermslii	Candidatus Hodgkinia cicadicola	Clostridium cellulovorans	Deinococcus proteolyticus
Acidithiobacillus ferrovorans	Arthrobacter arilaitensis	Borrelia recurrentis	Candidatus Korarchaeum cryptofilum	Clostridium kluyveri	Deinococcus radiodurans
Acidithiobacillus thiooxidans	Arthrobacter aurescens	Borrelia turicatae	Candidatus Koribacter versatilis	Clostridium lentocellum	Delftia acidovorans
Acidobacterium capsulatum	Arthrobacter chlorophenolicus	Brachyabacterium faecium	Candidatus Liberibacter asiaticus	Clostridium ljungdahlii	Delftia sp.
Acidobacterium sp.	Arthrobacter phenanthrenivorans	Brachyspira hyodysenteriae	Candidatus Liberibacter solanacearum	Clostridium novyi	Denitrovibrio acetiphilus
Acidothermus cellulolyticus	Arthrobacter sp.	Brachyspira murdochii	Candidatus Midichloria mitochondrii	Clostridium perfringens	Desulfarculus baarsii
Acidovorax avenae	Aster yellows witches'-broom phytoplasma"	Brachyspira pilosicoli	Candidatus Moranella endobia	Clostridium phytofermentans	Desulfatibacillum alkenivorans
Acidovorax citrullii	Asticcacaulis excentricus	Bradyrhizobium japonicum	Candidatus Nitrospira defluvi	Clostridium proteoclasticum	Desulfobacterium hafniense
Acidovorax ebreus	Atopobium parvulum	Bradyrhizobium sp.	Candidatus Pelagibacter ubique	Clostridium saccharolyticum	Desulfobacca acetoxidans
Acidovorax sp.	Azoarcus sp.	Brevibacillus brevis	Candidatus Protochlamydia amoebophila	Clostridium sp.	Desulfobacterium autotrophicum
Acidulipifundum boonei	Azorhizobium caulinodans	Brevundimonas subvibrioides	Candidatus Punicispirillum marinum	Clostridium tetani	Desulfobulbus propionicus
Acinetobacter baumannii	Azospirillum sp.	Brucella abortus	Candidatus Riesia pediculicola	Clostridium thermocellum	Desulfococcus oleovorans
Acinetobacter oleivorans	Azotobacter vinelandii	Brucella canis	Candidatus Ruthia magnifica	Collimonas fungivorans	Desulfococcus rebaense
Acinetobacter sp.	Bacillus amyloliquefaciens	Brucella melitensis	Candidatus Solibacter usitatus	Colwellia psychrerythraea	Desulfomicrobium baculatum
Actinobacillus pleuropneumoniae	Bacillus anthracis	Brucella microti	Candidatus Sulcia muelleri	Comamonas testosteroni	Desulfotalea psychrophila
Actinobacillus succinogenes	Bacillus atrophaeus	Brucella ovis	Candidatus Tremblaya princeps	Conexibacter woesei	Desulfotomaculum acetoxidans
Actinosynnema mirum	Bacillus cellulosilyticus	Burkholderia pinnipedialis	Candidatus Zindera insecticola	Coprothermobacter proteolyticus	Desulfotomaculum kuznetsovii
Aerococcus urinae	Bacillus cereus	Burkholderia suis	Capnocytophaga canimorsus	Coraliomargarita akajimensis	Desulfotomaculum reducens
Aeromonas hydrophila	Bacillus clausii	Buchnera aphidicola	Carboxydotherrhus hydrogenoformans	Coriobacterium glomerans	Desulfotomaculum ruminis
Aeromonas salmonicida	Bacillus coagulans	Burkholderia ambifaria	Carnobacterium sp.	Corynebacterium aurimucosum	Desulfotomaculum aespoensis
Aeromonas veronii	Bacillus cytotoxicus	Burkholderia cenocepacia	Catenulispora acidiphila	Corynebacterium diphtheriae	Desulfotomaculum alaskensis
Aeropyrum pernix	Bacillus halodurans	Burkholderia gladioli	Caulobacter segnis	Corynebacterium efficiens	Desulfotomaculum glutamicum
Aggregatibacter actinomycetemcomitans	Bacillus licheniformis	Burkholderia glumae	Caulobacter sp.	Corynebacterium glutamicum	Desulfotomaculum mallei
Aggregatibacter aphrophilus	Bacillus megaterium	Burkholderia mallei	Caulobacter sp.	Corynebacterium jeikeium	Desulfotomaculum mallei
Agrobacterium sp.	Bacillus pseudofirmus	Burkholderia multivorans	Caulobacter vibrioides	Corynebacterium kroppenstedtii	Desulfotomaculum salexigenis
Agrobacterium tumefaciens	Bacillus pumilus	Burkholderia phymatum	Cellulomonas fimi	Corynebacterium pseudotuberculosis	Desulfotomaculum vulgare
Agrobacterium vitis	Bacillus selenitireducens	Burkholderia phytofirmans	Cellulomonas flavigena	Corynebacterium resistens	Desulfurispirillum indicum
Akkermansia muciniphila	Bacillus subtilis	Burkholderia pseudomallei	Cellulophaga algicola	Corynebacterium ulcerans	Desulfurispirillum alkaliphilum
Albidiferax ferrireducens	Bacillus thuringiensis	Burkholderia rhizoxinica	Cellulophaga lytica	Corynebacterium urealyticum	Desulfurispirillum thermolithotrophum
Alcanivorax borkumensis	Bacillus weihenstephanensis	Burkholderia sp.	Cellvibrion japonicus	Corynebacterium variabile	Desulfurococcus kamchatkensis
Alicyclophilius denitrificans	Bacteroides fragilis	Burkholderia thailandensis	Cenarchaeum symbiosum	Coxiella burnetii	Desulfurococcus mucosus
Alicyclobacillus acidocaldarius	Bacteroides helcogenes	Burkholderia vietnamiensis	Chitinophaga pinensis	Croceibacter atlanticus	Dichelobacter nodosus
Allivibrio fischeri	Bacteroides salanitronis	Burkholderia xenovorans	Chlamydia muridarum	Cronobacter sakazakii	Dickeya dadantii
Allivibrio salmonicida	Bacteroides thetaiotaomicron	Caldanaerobacter subterraneus	Chlamydia pecorum	Cronobacter turicensis	Dickeya zeae
Alkalilimnicola ehrlichii	Bacteroides vulgatus	Caldicellulosiruptor bescii	Chlamydia pneumoniae	Cryptobacterium curtum	Desulfurococcus thermophilum
Alkaliphilus metallireducens	Bartonella bacilliformis	Caldicellulosiruptor hydrothermalis	Chlamydia psittaci	Cupriavidus metallidurans	Desulfurococcus turgidum
Alkaliphilus oremlandii	Bartonella clarridgeae	Caldicellulosiruptor kristjanssonii	Chlamydia trachomatis	Cupriavidus necator	Dinoroseobacter shibae
Allochromatium vinosum	Bartonella grahamii	Caldicellulosiruptor kronotskyensis	Chlamydia abortus		Dyadobacter fermentans
Alteromonas macleodii	Bartonella henselae	Caldicellulosiruptor lactoaceticus	Chlamydia caviae		Edwardsiella ictaluri
Alteromonas sp.	Bartonella quintana	Caldicellulosiruptor obsidiansis	Chlamydia felis		Edwardsiella tarda
Aminobacterium colombiense	Bartonella tribocorum	Caldicellulosiruptor owensensis	Chlorobaculum parvum		Eggerthella lenta
Ammonifex degensii	Baumannia cicadellincola	Caldicellulosiruptor saccharolyticus	Chlorobaculum tepidum		Eggerthella sp.
Amycolatopsis mediterranei	Bdellovibrio bacteriovorus	Calditerrivibrio nitroreducens	Chlorobium chlorochromatii		Ehrlichia canis
Amycolicoccus subflavus	Beijerinckia indica	Caldivirga maquilingensis	Chlorobium limicola		Ehrlichia chaffeensis
Anabaena variabilis	Beutenbergia cavernae	Caldyptogena okutanii thioautotrophic gill symbiont	Chlorobium phaeobacteroides		Ehrlichia ruminantium
Anaerococcus prevotii	Bifidobacterium adolescentis	Campylobacter concisus	Chlorobium phaeovibrioides		Elusimicrobium minutum

表一、原核生物序列特徵資料庫所包含的分類層級(續)

Species					
Polaromonas sp.	Rhodospirillum rubrum	Sideroxydans lithotrophicus	Sulfobacillus acidophilus	Thermosipho melanesiensis	Xanthomonas citri
Polymorphum gilvum	Rhodothermus marinus	Simkania negevensis	Sulfolobus acidocaldarius	Thermosphaera aggregans	Xanthomonas euvesicatoria
Polynucleobacter necessarius	Rickettsia africae	Sinorhizobium fredii	Sulfolobus islandicus	Thermosynechococcus elongatus	Xanthomonas oryzae
Porphyromonas asaccharolytica	Rickettsia akari	Sinorhizobium medicae	Sulfolobus solfataricus	Thermotoga lettingae	Xenorhabdus bovienii
Porphyromonas gingivalis	Rickettsia bellii	Sinorhizobium meliloti	Sulfolobus tokodaii	Thermotoga maritima	Xenorhabdus nematophila
Prevotella denticola	Rickettsia canadensis	Slackia heliotrinireducens	Sulfuricurvum kujiense	Thermotoga naphthophila	Xylanimonas cellulolytica
Prevotella melaninogenica	Rickettsia conorii	Sodalis glossinidius	Sulfurihydrogenibium azorense	Thermotoga neapolitana	Xylella fastidiosa
Prevotella ruminicola	Rickettsia felis	Sorangium cellulosum	Sulfurihydrogenibium sp.	Thermotoga petrophila	Yersinia enterocolitica
Prochlorococcus marinus	Rickettsia heilongjiangensis	Sphaerobacter thermophilus	Sulfurimonas autotrophica	Thermotoga sp.	Yersinia pestis
Propionibacterium acnes	Rickettsia massillae	Sphaerochaeta globus	Sulfurimonas denitrificans	Thermotoga thermarum	Yersinia pseudotuberculosis
Propionibacterium freudenreichii	Rickettsia peacockii	Sphingobacterium sp.	Sulfurospirillum deleyianum	Thermovibrio ammonificans	Zobellia galactanivorans
Prosthecochloris aestuarii	Rickettsia prowazekii	Sphingobium chlorophenolicum	Sulfurovum sp.	Thermus scotoductus	Zunongwangia profunda
Proteus mirabilis	Rickettsia rickettsii	Sphingobium japonicum	Symbiobacterium thermophilum	Thermus thermophilus	Zymomonas mobilis
Pseudoalteromonas atlantica	Rickettsia typhi	Sphingomonas wittichii	Syngochococcus elongatus	Thioalkalimicrobium cyclicum	
Pseudoalteromonas haloplanktis	Riemerella anatipestifer	Sphingopyxis alaskensis	Synechococcus sp.	Thioalkalivibrio sp.	
Pseudoalteromonas sp.	Robiginitalea biformata	Spirochaeta caldaria	Spirochaetocystis sp.	Thioalkalivibrio sulfidophilus	
Pseudomonas aeruginosa	Roseiflexus castenholzii	Spirochaeta coccoides	Syntrophobacter fumaroxidans	Thiobacillus denitrificans	
Pseudomonas brassicaearum	Roseiflexus sp.	Spirochaeta smaragdinae	Syntrophobacterium glycolicum	Thiomicrospira crunigena	
Pseudomonas entomophila	Roseobacter denitrificans	Spirochaeta thermophila	Syntrophomonas wolfei	Thiomonas intermedia	
Pseudomonas fluorescens	Roseobacter litoralis	Spirosoma linguale	Syntrophothermus lipocalidus	Tolomonas auensis	
Pseudomonas fulva	Rothia dentocariosa	Stackebrandtia nassauensis	Syntrophus aciditrophicus	Treponema azotonutricium	
Pseudomonas mendocina	Rothia mucilaginoso	Staphylococcus aureus	Staphylococcus aureus	Treponema brennaborensis	
Pseudomonas putida	Rubrobacter xylanophilus	Staphylococcus carnosus	Staphylococcus carnosus	Treponema denticola	
Pseudomonas savastanoi	Ruegeria pomeroyi	Staphylococcus epidermidis	Staphylococcus epidermidis	Treponema pallidum	
Pseudomonas stutzeri	Ruegeria sp.	Staphylococcus haemolyticus	Staphylococcus haemolyticus	Treponema paraluis-cuniculi	
Pseudomonas syringae	Ruminococcus albus	Staphylococcus lugdunensis	Staphylococcus lugdunensis	Treponema primitia	
Pseudomonas syringae group	Runella sliathyformis	Staphylococcus pseudintermedius	Staphylococcus pseudintermedius	Treponema succinifaciens	
genomosa 3					
Pseudonocardia dioxanivorans	Saccharomonospora viridis	Staphylococcus saprophyticus	Staphylococcus saprophyticus	Trichodesmium erythraeum	
Pseudoxanthomonas suwonensis	Saccharophagus degradans	Staphylothermus hellenicus	Staphylothermus hellenicus	Trichormus azollae	
Psychrobacter arcticus	Saccharopolyspora erythraea	Staphylothermus marinus	Staphylothermus marinus	Tropheryma whipplei	
Psychrobacter cryohalolentis	Salinibacter ruber	Starkeya novella	Starkeya novella	Truepera radiovictrix	
Psychrobacter sp.	Salinispora arenicola	Stenotrophomonas maltophilia	Stenotrophomonas maltophilia	Tsakamurella paurometabola	
Psychromonas ingrahamii	Salinispora tropica	Stigmatella aurantiaca	Stigmatella aurantiaca	Uncultured methanogenic archaeon RC-1	
Pusillimonas sp.	Salmonella bongori	Streptobacillus moniliformis	Streptobacillus moniliformis	uncultured Termite group 1 bacterium	
Pyrobaculum aerophilum	Salmonella enterica	Streptococcus agalactiae	Streptococcus agalactiae	Ureaplasma parvum	
Pyrobaculum arsenaticum	Sanguibacter keddieii	Streptococcus dysgalactiae	Streptococcus dysgalactiae	Ureaplasma urealyticum	
Pyrobaculum caldifontis	Sebaldella termitidis	Streptococcus equi	Streptococcus equi	Variovorax paradoxus	
Pyrobaculum islandicum	Segniliparus rotundus	Streptococcus gallolyticus	Streptococcus gallolyticus	Variolona parvula	
Pyrococcus abyssi	Selenomonas sputigena	Streptococcus gordonii	Streptococcus gordonii	Verminephrobacter eiseniae	
Pyrococcus furiosus	Serratia plymuthica	Streptococcus mitis	Streptococcus mitis	Verrucosipora maris	
Pyrococcus horikoshii	Serratia proteamaculans	Streptococcus mutans	Streptococcus mutans	Vibrio cholerae	
Pyrococcus sp.	Serratia sp.	Streptococcus oralis	Streptococcus oralis	Vibrio cholerae	
Pyrococcus yayanosii	Shewanella amazonensis	Streptococcus parasanguinis	Streptococcus parasanguinis	Vibrio harveyi	
Rahnella sp.	Shewanella baltica	Streptococcus parauberis	Streptococcus parauberis	Vibrio parahaemolyticus	
Ralstonia pickettii	Shewanella denitrificans	Streptococcus pasteurianus	Streptococcus pasteurianus	Vibrio sp.	
Ralstonia solanacearum	Shewanella frigidimarina	Streptococcus pneumoniae	Streptococcus pneumoniae	Vibrio splendidus	
Ramlibacter tataouinensis	Shewanella halifaxensis	Streptococcus pseudopneumoniae	Streptococcus pseudopneumoniae	Vibrio vulnificus	
Renibacterium salmoninarum	Shewanella loihica	Streptococcus pyogenes	Streptococcus pyogenes	Vulcanisaeta distributa	
Rhizobium etli	Shewanella oneidensis	Streptococcus salivarius	Streptococcus salivarius	Vulcanisaeta moutnovskia	
Rhizobium leguminosarum	Shewanella pealeana	Streptococcus sanguinis	Streptococcus sanguinis	Waddlia chondrophila	
Rhodobacter capsulatus	Shewanella piezotolerans	Streptococcus suis	Streptococcus suis	Weeksella virosa	
Rhodobacter sphaeroides	Shewanella putrefaciens	Streptococcus thermophilus	Streptococcus thermophilus	Weissella koreensis	
Rhodococcus equi	Shewanella sediminis	Streptococcus uberis	Streptococcus uberis	Wigglesworthia glossinidia	
Rhodococcus erythropolis	Shewanella sp.	Streptomyces avermitilis	Streptomyces avermitilis	Wolbachia endosymbiont of Brugia malayi	
Rhodococcus jostii	Shewanella violacea	Streptomyces coelicolor	Streptomyces coelicolor	Wolbachia endosymbiont of Culex quinquefasciatus	
Rhodococcus opacus	Shewanella woodyi	Streptomyces griseus	Streptomyces griseus	Wolbachia endosymbiont of Drosophila melanogaster	
Rhodomicrobium vannielii	Shigella boydii	Streptomyces scabiei	Streptomyces scabiei	Wolbachia sp.	
Rhodopirellula baltica	Shigella dysenteriae	Streptomyces sp.	Streptomyces sp.	Wolinella succinigenes	
Rhodopseudomonas palustris	Shigella flexneri	Streptomyces violaceiniger	Streptomyces violaceiniger	Xanthobacter autotrophicus	
Rhodospirillum centenum	Shigella sonnei	Streptosporangium roseum	Streptosporangium roseum	Xanthomonas albilineans	
				Xanthomonas campestris	

表二、真核生物序列特徵資料庫所包含的分類層級

Kingdom	Phylum	Class	Order	Family	Genus	Species
Metazoa	Arthropoda	Tremellomycetes	Mamiellales	Saccharomycetaceae	Filobasidiella	Kluyveromyces thermotolerans
Fungi	Microsporidia	Liliopsida	Cryptomonadales	Cryptomonadaceae	Magnaporthe	Caenorhabditis briggsae
Viridiplantae	Apicomplexa	Schizosaccharomycetes	Rhabditida	Geminigeraceae	Zygosaccharomyces	Candida albicans
	Nematoda	Mamiellophyceae	Pyrenomonadales	Phaeodactylaceae	Yarrowia	Scheffersomyces stipitis
	Ascomycota	Chromadorea	Schizosaccharomycetales	Poaceae	Arabidopsis	Encephalitozoon cuniculi
	Basidiomycota	Oligohymenophorea	Poales	Cyanophoraceae	Cyanophora	Trypanosoma brucei
	Chlorophyta	Cryptophyta	Haemosporida	Parameciidae	Plasmodium	Plasmodium knowlesi
	Bacillariophyta	Saccharomycetes	Diptera	Brassicaceae	Oryza	Candida dubliniensis
	Streptophyta	Glaucozystophyceae	Saccharomycetales	Babesiidae	Caenorhabditis	Theileria parva
		Coscinodiscophyceae	Peniculida	Magnaporthaceae	Cryptomonas	Candida glabrata
		Bacillariophyceae	Magnaporthales	Trypanosomatidae	Schizosaccharomyces	Thalassiosira pseudonana
		Insecta	Naviculales	Debaryomycetaceae	Paramecium	Phaeodactylum tricornutum
		Aconoidasida	Tremellales	Dipodascaceae	Thalassiosira	Cryptococcus neoformans
		Sordariomycetes	Kinetoplastida	Thalassiosiraceae	Candida	Guillardia theta
			Brassicales	Tremellaceae	Guillardia	Leishmania major
				Drosophilidae	Kluyveromyces	Plasmodium falciparum
				Hemiselmidaceae	Theileria	Debaryomyces hansenii
				Rhabditidae	Debaryomyces	Cryptomonas paramecium
				Theileriidae	Eremothecium	Cryptococcus gattii
				Unikaryonidae	Micromonas	Eremothecium gossypii
				Schizosaccharomycetaceae	Babesia	Magnaporthe oryzae
					Hemiselmis	Caenorhabditis elegans
					Trypanosoma	Komagataella pastoris
					Phaeodactylum	Paramecium tetraurelia
					Bigelowiella	Hemiselmis andersenii
					Lachancea	Ostreococcus 'lucimarinus'
					Drosophila	Zygosaccharomyces rouxii
					Nakaseomyces	Yarrowia lipolytica
					Encephalitozoon	Babesia bovis
					Saccharomyces	Cyanophora paradoxa
					Ostreococcus	Encephalitozoon intestinalis
					Komagataella	Bigelowiella natans
					Scheffersomyces	Schizosaccharomyces pombe
					Oryza sativa	Oryza sativa
					Micromonas sp. RCC299	Micromonas sp. RCC299
					Arabidopsis thaliana	Arabidopsis thaliana
					Kluyveromyces lactis	Kluyveromyces lactis
					Saccharomyces cerevisiae	Saccharomyces cerevisiae
					Leishmania infantum	Leishmania infantum

表三、測試序列的物種名稱

二十個不同strain的人體腸道細菌染色體序列	三十種人體腸道細菌染色體序列	三十種真核生物測試序列	三十個種層級不存於原核資料庫的測試序列	三十個屬層級不存於原核資料庫的測試序列
Bacteroides fragilis NCTC 9343	Bacteroides fragilis NCTC9343	Candida albicans	Acinetobacter calcoaceticus	Anaerobaculum hydrogeniformans
Bacteroides fragilis YCH46	Bacteroides helcogene P36-108	Candida dubliniensis	Acinetobacter radioresistens	Algoriphagus sp. PR1
Bacteroides helcogenes P 36-108	Bacteroides salanitronis DSM18170	Candida glabrata	Aggregatibacter segnis	Actinomyces sp.
Bacteroides salanitronis DSM 18170	Bacteroides thetaiotaomicron VPI5482	Cryptococcus gattii	Anaerococcus vaginalis	Actinomyces urogenitalis
Bacteroides thetaiotaomicron VPI-5482	Bacteroides vulgatus ATCC8482	Debaryomyces hansenii	Brucella ceti	Centipeda periodontii
Bacteroides vulgatus ATCC 8482	*Clostridium acetobutylicum DSM1731	Encephalitozoon cuniculi	Burkholderia dolosa	Dokdonia donghaensis
Prevotella denticola F0289	*Clostridium botulinum BKT015925	Encephalitozoon intestinalis	Brucella neotomae	Dialister invisus
Prevotella melaninogenica ATCC 25845 chromosome chromosome I	*Clostridium difficile BI1	Magnaporthe oryzae	Enterobacter cancerogenus	Jonquetella anthropi
Prevotella ruminicola 23 chromosome	*Clostridium sp. SY8519	Pichia pastoris	Enterococcus casseliflavus	Janibacter sp.
Parabacteroides distasonis ATCC 8503	Escherichia coli 536	Saccharomyces cerevisiae	Enterococcus gallinarum	Kordia algicida
Lactobacillus delbrueckii 2038	*Eubacterium eligens ATCC27750	Schizosaccharomyces pombe	Eubacterium saphenum	Kingella denitrificans
Lactobacillus fermentum IFO 3956	*Eubacterium rectale ATCC33656	Yarrowia lipolytica	Fusobacterium varium	Labrenzia alexandrii
Lactococcus garvieae Lg2	*Lactobacillus crispatus ST1	Caenorhabditis briggsae	Helicobacter canadensis	Leeuwenhoekiiella blandensis
Lactococcus lactis subsp. lactis KF147	Lactobacillus delbrueckii 2038	Caenorhabditis elegans	Lactobacillus iners	Loktanella vestfoldensis
Streptococcus gallolyticus BAA-2069	Lactobacillus fermentum IFO3956	Drosophila melanogaster	Listeria ivanovii	Maritimibacter alkaliphilus
Streptococcus salivarius CCHSS3	Lactococcus lactis KF147	Arabidopsis thaliana	Mycobacterium kansasii	Roseibium sp.
Escherichia coli 536	*Leuconostoc citreum KM20	Micromonas sp.	Pediococcus acidilactici	Methylobacter tundripaludum
Escherichia coli APEC O1	*Methanobrevibacter smithii ATCC35061	Oryza sativa Japonica Group DNA	Prevotella bergensis	Oceanicaulis alexandrii
Escherichia coli NA114	*Methanosphaera stadtmanae DSM3091	Ostreococcus lucimarinus	Prevotella buccae	Octadecabacter antarcticus
Escherichia coli UMNK88	Parabacteroides distasonis ATCC8503	Babesia bovis	Parabacteroides johnsonii	Oxalobacter formigenes
	Prevotella denticola F0289	Cryptomonas paramecium	Streptomyces albus	Sulfitobacter sp.
	Prevotella melaninogenica ATCC25845	Guillardia theta	Streptomyces clavuligerus	Pseudoramibacter alactolyticus
	Prevotella ruminicola 23	Hemiselmis andersenii	Streptococcus equinus	Polaribacter irgensii
	*Ruminococcus albus 7	Leishmania infantum	Streptomyces griseoflavus	Providencia rustigianii
	*Streptococcus equi subsp zoepidemicus MGCS10565	Leishmania major	Streptomyces lividans	Peptoniphilus sp.
	Streptococcus gallolyticus BAA2069	Paramecium tetraurelia	Serratia odorifera	Psychroflexus torquis
	*Streptococcus mitis B6	Phaeodactylum tricornutum	Segniliparus rugosus	Roseovarius nubinhibens
	Streptococcus salivarius CCHSS3	Plasmodium knowlesi	Sphingobacterium spiritivorum	Roseovarius sp.
	*Streptococcus suis SC84	Thalassiosira pseudonana	Vibrio alginolyticus	Scardovia inopinata
	Lactococcus garvieae Lg2	Trypanosoma brucei	Veillonella dispar	Subdoligranulum variabile

*代表新增加的十四種人體腸道細菌染色體序列

表四、*Prevotella denticola* 在各分類階層的分佈情形

Phylum	Percentag	Class	Percentag	Order	Percentag	Family	Percentag	Genus	Percentag
Chlorobi	52.94%	Chlorobia	52.94%	Chlorobiales	52.94%	Chlorobiaceae	52.94%	Pelodictyon	47.64%
Bacteroidetes	16.72%	Bacteroidia	16.46%	Bacteroidales	16.46%	Bacteroidaceae	9.54%	Bacteroides	9.54%
Proteobacteria	11.38%	Clostridia	6.56%	Clostridiales	5.80%	Porphyromonadaceae	6.54%	Porphyromonas	6.52%
Firmicutes	10.86%	Deltaproteobacteria	4.96%	Bacillales	2.98%	Clostridiaceae	2.42%	Chlorobium	2.76%
Euryarchaeota	4.40%	Bacilli	4.30%	Desulfuromonadales	2.20%	Eubacteriaceae	2.24%	Prosthecochloris	2.46%
Thermotogae	1.18%	Methanomicrobia	2.68%	Methanosarcinales	1.82%	Geobacteraceae	2.06%	Eubacterium	2.24%
Actinobacteria	0.96%	Gammaproteobacteria	2.24%	Syntrophobacterales	1.36%	Bacillaceae	2.04%	Clostridium	2.16%
Cyanobacteria	0.44%	Epsilonproteobacteria	1.96%	Lactobacillales	1.32%	Methanosarcinaceae	1.76%	Geobacter	2.06%
Spirochaetes	0.30%	Alphaproteobacteria	1.32%	Thermotogales	1.18%	Syntrophaceae	1.34%	Sulfurovum	1.82%
Verrucomicrobia	0.18%	Thermotogae	1.18%	Thermoplasmatales	1.02%	Thermotogaceae	1.18%	Bacillus	1.68%
Chlamydiae	0.16%	Thermoplasmata	1.02%	Rickettsiales	0.90%	Streptococcaceae	1.16%	Syntrophus	1.34%
Crenarchaeota	0.16%	Actinobacteria	0.96%	Methanomicrobiales	0.86%	Thermoplasmataceae	1.02%	Methanosarcina	1.16%
Chloroflexi	0.10%	Betaproteobacteria	0.90%	Thermoanaerobacterales	0.70%	Peptococcaceae	0.82%	Streptococcus	1.16%
Acidobacteria	0.08%	Archaeoglobi	0.26%	Actinomycetales	0.68%	Nitrosomonadaceae	0.68%	Thermoplasma	1.02%
Elusimicrobia	0.06%	Thermococci	0.20%	Nitrosomonadales	0.68%	Anaplasmataceae	0.64%	Methanococcoides	0.60%
Aquificae	0.06%	Halobacteria	0.18%	Desulfuovibrionales	0.64%	Desulfuovibrionaceae	0.64%	Desulfuovibrio	0.56%
Nitrospirae	0.02%	Thermoprotei	0.16%	Alteromonadales	0.56%	Thermodesulfobiaceae	0.42%	Desulfotobacterium	0.50%
		Verrucomicrobiae	0.16%	Desulfobacterales	0.52%	Desulfobulbaceae	0.40%	Nitrosomonas	0.50%
		Aquificae	0.06%	Enterobacterales	0.36%	Staphylococcaceae	0.38%	Kosmotoga	0.46%
		Elusimicrobia	0.06%	Spirochaetales	0.30%	Enterobacteriaceae	0.36%	Neorickettsia	0.46%
		Dehalococcoidetes	0.04%	Bifidobacteriales	0.28%	Methanospirillaceae	0.36%	Coprothermobacter	0.42%
		Chloroflexi	0.04%	Rhizobiales	0.28%	Methanocorpusculaceae	0.34%	Desulfotalea	0.40%
		Methanopyri	0.04%	Vibrionales	0.28%	Corynebacteriaceae	0.32%	Candidatus Azobacteroides	0.38%
		Nitrospira	0.02%	Archaeoglobales	0.26%	Shewanellaceae	0.30%	Methanospirillum	0.36%
		Methanococci	0.02%	Legionellales	0.26%	Spirochaetaeaceae	0.30%	Methanocorpusculum	0.34%
		Thermomicrobia	0.02%	Prochlorales	0.22%	Bifidobacteriaceae	0.28%	Thermotoga	0.34%
				Xanthomonadales	0.22%	Paenibacillaceae	0.28%	Corynebacterium	0.32%
				Thermococcales	0.20%	Vibrionaceae	0.28%	Shewanella	0.30%
				Bdellovibrionales	0.18%	Archaeoglobaceae	0.26%	Bifidobacterium	0.28%
				Halobacteriales	0.18%	Rickettsiaceae	0.24%	Brevibacillus	0.28%
				Verrucomicrobiales	0.16%	Legionellaceae	0.22%	Exiguobacterium	0.28%
				Chlamydiales	0.16%	Prochlorococcaceae	0.22%	Fervidobacterium	0.28%
				Sphingobacteriales	0.14%	Xanthomonadaceae	0.22%	Vibrio	0.28%
				Burkholderiales	0.14%	Thermococcaceae	0.20%	Archaeoglobus	0.26%
				Flavobacteriales	0.12%	Bdellovibrionaceae	0.18%	Treponema	0.24%
				Sulfolobales	0.10%	Halobacteriaceae	0.18%	Legionella	0.22%
				Pasteurellales	0.10%	Thermoanaerobacteraceae	0.18%	Prochlorococcus	0.22%
				Thiotrichales	0.10%	Hellobacteriaceae	0.16%	Xylella	0.22%
				Oceanospirillales	0.10%	Verrucomicrobiaceae	0.16%	Macrococcus	0.20%
				Nostocales	0.10%	Chlamydiaceae	0.14%	Oceanobacillus	0.20%
				Rhodospirillales	0.08%	Pelobacteraceae	0.14%	Pelotomaculum	0.20%
				Chroococcales	0.08%	Alteromonadaceae	0.12%	Thermococcus	0.20%
				Pseudomonadales	0.08%	Desulfobacteraceae	0.12%	Bdellovibrio	0.18%
				Myxococcales	0.06%	Flavobacteriaceae	0.12%	Nitrosospira	0.18%
				Aquificales	0.06%	Bartonellaceae	0.10%	Staphylococcus	0.18%
				Desulfurococcales	0.06%	Burkholderiaceae	0.10%	Trophenyia	0.18%
				Methylophilales	0.06%	Nostocaceae	0.10%	Akkermansia	0.16%
				Natranaerobiales	0.06%	Pasteurellaceae	0.10%	Anaplasma	0.16%
				Elusimicrobiales	0.06%	Rhizobiaceae	0.10%	Haloaquadratum	0.16%
				Methylococcales	0.06%	Sulfolobaceae	0.10%	Hellobacterium	0.16%
				Methanopyrales	0.04%	Syntrophomonadaceae	0.10%	Wolbachia	0.16%
				Sphingomonadales	0.04%	Thermoanaerobacterales	0.10%	Alkaliphilus	0.14%
				Acidobacteriales	0.04%	Acetobacteraceae	0.08%	Chlamydothila	0.14%
				Chloroflexales	0.04%	Idiomarinaceae	0.08%	Pelobacter	0.14%
				Chromatiales	0.04%	Pseudomonadaceae	0.08%	Thermoanaerobacter	0.14%
				Campylobacteriales	0.04%	Acidothermaceae	0.06%	Methanosphaerula	0.12%
				Nitrospirales	0.02%	Clostridiales	0.06%	Bartonella	0.10%
				Rhodobacterales	0.02%	Elusimicrobiaceae	0.06%	Caldicellulosiruptor	0.10%
				Acidithiobacillales	0.02%	Lactobacillaceae	0.06%	Cytophaga	0.10%
				Nautiliales	0.02%	Leuconostocaceae	0.06%	Desulfotomaculum	0.10%
				Hydrogenophilales	0.02%	Methanosphaetaceae	0.06%	Flavobacterium	0.10%
				Methylacidiphilales	0.02%	Methylococcaceae	0.06%	Marinobacter	0.10%
				Methanococcales	0.02%	Methylophilaceae	0.06%	Nostoc	0.10%
				Oscillatoriales	0.02%	Natranaerobiaceae	0.06%	Sulfobolus	0.10%
				Thermomicrobiales	0.02%	Nocardiaceae	0.06%	Syntrophomonas	0.10%
						Piscinickettsiaceae	0.06%	Desulfobacterium	0.08%
						Psychromonadaceae	0.06%	Geobacillus	0.08%
						Acidobacteriaceae	0.04%	Haemophilus	0.08%
						Alcanivoracaceae	0.04%	Idiomarina	0.08%
						Aquificaceae	0.04%	Lawsonia	0.08%
						Bruceellaceae	0.04%	Lysinibacillus	0.08%
						Chloroflexaceae	0.04%	Nitratiruptor	0.08%
						Chromatiaceae	0.04%	Photorhabdus	0.08%
						Coxiellaceae	0.04%	Pseudomonas	0.08%
						Desulfurococccaceae	0.04%	Acidothemus	0.06%
						Enterococccaceae	0.04%	Borrelia	0.06%
						Francisellaceae	0.04%	Chlorobaculum	0.06%
						Halomonadaceae	0.04%	Elusimicrobium	0.06%
						Helicobacteraceae	0.04%	Fingoldia	0.06%
						Methanopyraceae	0.04%	Granulibacter	0.06%
						Mycobacteriaceae	0.04%	Lactobacillus	0.06%
						Myxococcaceae	0.04%	Methanoseta	0.06%
						Rhodothermaceae	0.04%	Methylobacillus	0.06%
						Acidithiobacillaceae	0.02%	Methylococcus	0.06%

表四、*Prevotella denticola* 在各分類階層的分類情形(續)

Phylum	Percentag Class	Percentag Order	Percentag Family	Percentag Genus	Percentag	
			Erythrobacteraceae	0.02%	Natranaerobius	0.06%
			Hahellaceae	0.02%	Pectobacterium	0.06%
			Hydrogenophilaceae	0.02%	Psychromonas	0.06%
			Hydrogenothermaceae	0.02%	Ralstonia	0.06%
			Methanococcaceae	0.02%	Rhodococcus	0.06%
			Methylacidiphilaceae	0.02%	Rickettsia	0.06%
			Nautiliaceae	0.02%	Shigella	0.06%
			Nitrospiraceae	0.02%	Synechococcus	0.06%
			Oxalobacteraceae	0.02%	Thermosiphon	0.06%
			Parachlamydiaceae	0.02%	Thiomicrospira	0.06%
			Phyllobacteriaceae	0.02%	Acidobacterium	0.04%
			Polyangiaceae	0.02%	Alcanivorax	0.04%
			Pyrodictiaceae	0.02%	Candidatus Hamiltonella	0.04%
			Rhodobacteraceae	0.02%	Candidatus Koribacter	0.04%
			Sphingomonadaceae	0.02%	Chloroflexus	0.04%
			Streptomycetaceae	0.02%	Chromohalobacter	0.04%
			Syntrophobacteraceae	0.02%	Coxiella	0.04%
			Thermomicrobiaceae	0.02%	Cupriavidus	0.04%
			Xanthobacteraceae	0.02%	Dehalococcoides	0.04%
					Desulfococcus	0.04%
					Desulfurococcus	0.04%
					Enterococcus	0.04%
					Francisella	0.04%
					Helicobacter	0.04%
					Hydrogenobaculum	0.04%
					Leuconostoc	0.04%
					Methanopyrus	0.04%
					Moorella	0.04%
					Mycobacterium	0.04%
					Myxococcus	0.04%
					Nitrosococcus	0.04%
					Ochrobactrum	0.04%
					Petrogoga	0.04%
					Rhizobium	0.04%
					Salinibacter	0.04%
					Sinorhizobium	0.04%
					Yersinia	0.04%
					Acaryochloris	0.02%
					Acidithiobacillus	0.02%
					Agrobacterium	0.02%
					Alteromonas	0.02%
					Buchnera	0.02%
					Candidatus Blochmannia	0.02%
					Candidatus Desulfonidis	0.02%
					Candidatus Pelagibacter	0.02%
					Candidatus Protochlamydia	0.02%
					Chlorohepeton	0.02%
					Cyanothece	0.02%
					Ehrlichia	0.02%
					Erwinia	0.02%
					Erythrobacter	0.02%
					Gluconobacter	0.02%
					Gramella	0.02%
					Hahella	0.02%
					Halobacterium	0.02%
					Hyperthermus	0.02%
					Janthinobacterium	0.02%
					Klebsiella	0.02%
					Mesorhizobium	0.02%
					Methanococcus	0.02%
					Methylacidiphilum	0.02%
					Methylbium	0.02%
					Nautilia	0.02%
					Novosphingobium	0.02%
					Oenococcus	0.02%
					Orientia	0.02%
					Parabacteroides	0.02%
					Pasteurella	0.02%
					Roseobacter	0.02%
					Sorangium	0.02%
					Streptomyces	0.02%
					Sulfurihydrogenibium	0.02%
					Syntrophobacter	0.02%
					Thermodesulfobivrio	0.02%
					Thermomicrobium	0.02%
					Thiobacillus	0.02%
					Trichodesmium	0.02%
					Xanthobacter	0.02%

表五、Prevotella ruminicola 在各分類階層的分類情形

Phylum	Percentag	Class	Percentag	Order	Percentag	Family	Percentag	Genus	Percentag
Proteobacteria	56.16%	Gammaproteobacteria	42.62%	Alteromonadales	24.84%	Shewanellaceae	20.54%	Shewanella	20.54%
Firmicutes	12.36%	Bacteroidia	7.50%	Bacteroidales	7.50%	Enterobacteriaceae	5.50%	Bacteroides	5.06%
Bacteroidetes	8.22%	Bacilli	6.98%	Enterobacteriales	5.50%	Bacteroidaceae	5.06%	Vibrio	3.66%
Euryarchaeota	7.28%	Clostridia	5.38%	Chlorobiales	5.06%	Chlorobiaceae	5.06%	Corynebacterium	2.66%
Chlorobi	5.06%	Chlorobia	5.06%	Actinomycetales	4.48%	Vibrionaceae	4.26%	Nitrosomonas	2.66%
Actinobacteria	4.76%	Actinobacteria	4.76%	Vibrionales	4.26%	Anaplasmataceae	3.44%	Anaplasma	2.52%
Cyanobacteria	2.12%	Betaproteobacteria	4.48%	Lactobacillales	4.24%	Alteromonadaceae	3.06%	Streptococcus	2.44%
Thermotogae	1.26%	Alphaproteobacteria	4.08%	Rickettsiales	3.56%	Nitrosomonadaceae	3.04%	Thermococcus	2.38%
Crenarchaeota	0.98%	Deltaproteobacteria	3.88%	Clostridiales	3.50%	Corynebacteriaceae	2.66%	Pelodictyon	2.32%
Chloroflexi	0.72%	Thermococci	2.54%	Nitrosomonadales	3.04%	Thermococcaceae	2.54%	Desulfotalea	2.02%
Chlamydiae	0.58%	Thermoplasmata	1.78%	Bacillales	2.74%	Streptococcaceae	2.44%	Porphyromonas	1.92%
Spirochaetes	0.28%	Methanomicrobia	1.56%	Thermococcales	2.54%	Desulfotribulaceae	2.02%	Thermoplasma	1.78%
Acidobacteria	0.12%	Thermotogae	1.26%	Pseudomonadales	2.22%	Porphyromonadaceae	2.00%	Coprothermobacter	1.58%
Verrucomicrobia	0.04%	Thermoprotei	0.98%	Desulfobacterales	2.04%	Thermoplasmataceae	1.78%	Xylella	1.52%
Nitrospirae	0.02%	Chloroflexi	0.70%	Thermoanaerobacterales	1.78%	Bacillaceae	1.72%	Marinobacter	1.44%
Thaumarchaeota	0.02%	Halobacteria	0.66%	Thermoplasmatales	1.78%	Clostridiaceae	1.62%	Yersinia	1.44%
		Archaeoglobi	0.58%	Desulfuromonadales	1.56%	Thermodesulfobiaceae	1.58%	Pectobacterium	1.26%
		Epsilonproteobacteria	0.54%	Xanthomonadales	1.52%	Xanthomonadaceae	1.52%	Prochlorococcus	1.22%
		Methanobacteria	0.06%	Oceanospirillales	1.44%	Thermotogaceae	1.26%	Chlorobium	1.12%
		Methanopyri	0.06%	Methanosarcinales	1.26%	Prochlorococcaceae	1.22%	Clostridium	1.04%
		Methanococci	0.04%	Thermotogales	1.26%	Pseudomonadaceae	1.18%	Alcanivorax	1.00%
		Dehalococcoidetes	0.02%	Legionellales	1.22%	Lactobacillaceae	1.12%	Alteromonas	1.00%
		Deinococci	0.02%	Prochlorales	1.22%	Moraxellaceae	1.04%	Lactobacillus	1.00%
		Nitrospira	0.02%	Methylophilales	0.88%	Alcanivoracaceae	1.00%	Mycobacterium	1.00%
				Desulfurococcales	0.66%	Mycobacteriaceae	1.00%	Geobacter	0.96%
				Halobacteriales	0.66%	Geobacteraceae	0.96%	Prosthecochloris	0.96%
				Chloroflexales	0.60%	Eubacteriaceae	0.94%	Eubacterium	0.94%
				Archaeoglobales	0.58%	Methylophilaceae	0.88%	Methylobacillus	0.88%
				Chlamydiales	0.58%	Legionellaceae	0.86%	Legionella	0.86%
				Aeromonadales	0.50%	Methanosarcinaceae	0.84%	Bacillus	0.82%
				Burkholderiales	0.50%	Peptococcaceae	0.70%	Neorickettsia	0.78%
				Chromatiales	0.50%	Halobacteriaceae	0.66%	Cellvibrio	0.72%
				Rhizobiales	0.48%	Staphylococcaceae	0.64%	Tropheryma	0.68%
				Flavobacteriales	0.32%	Leucosotocaceae	0.62%	Acinetobacter	0.64%
				Methanomicrobiales	0.30%	Chloroflexaceae	0.60%	Haloquadratum	0.64%
				Acidithiobacillales	0.28%	Pelobacteraceae	0.60%	Saccharophagus	0.62%
				Bifidobacteriales	0.28%	Pyrodictiaceae	0.60%	Chloroflexus	0.60%
				Spirochaetales	0.28%	Archaeoglobaceae	0.58%	Hyperthermus	0.60%
				Chroococcales	0.26%	Chlamydiaceae	0.54%	Pelobacter	0.60%
				Sulfolobales	0.24%	Aeromonadaceae	0.50%	Archaeoglobus	0.58%
				Sphingobacteriales	0.20%	Chromataceae	0.50%	Chlorobaculum	0.56%
				Thiotrichales	0.16%	Bartonellaceae	0.48%	Leuconostoc	0.56%
				Bdellovibrionales	0.14%	Idiomarinaceae	0.48%	Magnetococcus	0.56%
				Desulfovibrionales	0.14%	Methanosarcinaceae	0.42%	Escherichia	0.54%
				Nostocales	0.14%	Pseudoalteromonadaceae	0.38%	Alkaliphilus	0.52%
				Pasteurellales	0.12%	Coxiellaceae	0.36%	Methanococcoides	0.52%
				Herpetosiphonales	0.10%	Burkholderiaceae	0.32%	Photobacterium	0.52%
				Natranaerobiales	0.10%	Flavobacteriaceae	0.32%	Nitrosococcus	0.50%
				Thermoproteales	0.08%	Oceanospirillaceae	0.32%	Acanyochloris	0.48%
				Methanobacteriales	0.06%	Acidithiobacillaceae	0.28%	Bartonella	0.48%
				Methanopyrales	0.06%	Bifidobacteriaceae	0.28%	Idiomarina	0.48%
				Rhodocyclales	0.06%	Psychromonadaceae	0.28%	Sulfurovum	0.48%
				Methanococcales	0.04%	Spirochaetaceae	0.26%	Macrococcus	0.46%
				Methylacidiphilales	0.04%	Sulfobacillaceae	0.24%	Pseudomonas	0.46%
				Campylobacteriales	0.02%	Paenibacillaceae	0.20%	Candidatus Azobacteroides	0.44%
				Deinococcales	0.02%	Listeriaceae	0.18%	Erwinia	0.42%
				Nautiiales	0.02%	Oxalobacteraceae	0.16%	Methanosarcina	0.42%
				Nitrospumilales	0.02%	Thermoanaerobacteraceae	0.16%	Psychrobacter	0.40%
				Nitrospirales	0.02%	Bdellovibrionaceae	0.14%	Chlamydomyxa	0.38%
				Oscillatoriales	0.02%	Desulfovibrionaceae	0.14%	Edwardsiella	0.38%
				Rhodobacteriales	0.02%	Nostocaceae	0.14%	Lysinibacillus	0.38%
				Sphingomonadales	0.02%	Pasteurellaceae	0.12%	Nitrospira	0.38%
						Propionibacteriaceae	0.12%	Pseudoalteromonas	0.38%
						Rickettsiaceae	0.12%	Coxiella	0.36%
						Syntrophomonadaceae	0.12%	Desulfotomaculum	0.36%
						Clostridiales	0.10%	Kosmotoga	0.36%
						Colwelliaceae	0.10%	Oceanobacillus	0.34%
						Hahellaceae	0.10%	Ferribacterium	0.32%
						Herpetosiphonaceae	0.10%	Marinomonas	0.32%
						Natranaerobiaceae	0.10%	Methanosarcina	0.32%
						Piscirickettsiaceae	0.10%	Polynucleobacter	0.32%
						Desulfurococccaceae	0.06%	Desulfitobacterium	0.30%
						Enterococcaceae	0.06%	Acidithiobacillus	0.28%
						Francisellaceae	0.06%	Bifidobacterium	0.28%
						Methanobacteriaceae	0.06%	Petrogoga	0.28%
						Methanococcusulaceae	0.06%	Psychromonas	0.28%
						Methanopyraceae	0.06%	Thermotoga	0.28%
						Methanospirillaceae	0.06%	Aeromonas	0.26%
						Rhodocyclaceae	0.06%	Candidatus Hamiltonella	0.26%
						Methanococcaceae	0.04%	Shigella	0.24%
						Methylacidiphilaceae	0.04%	Tolomonas	0.24%
						Parachlamydiaceae	0.04%	Synechococcus	0.22%
						Thermoanaerobacterales	0.04%	Treponema	0.22%
						Thermofilaceae	0.04%	Brevibacillus	0.20%

表五、*Prevotella ruminicola* 在各分類階層的分類情形(續)

Phylum	Percentag Class	Percentag Order	Percentag Family	Percentag Genus	Percentag
			Thermoproteaceae	Candidatus Amoebophilus	0.20%
			Campylobacteraceae	Cytophaga	0.20%
			Comamonadaceae	Flavobacterium	0.20%
			Deinococcaceae	Dickeya	0.18%
			Desulfobacteraceae	Listeria	0.18%
			Halomonadaceae	Serratia	0.18%
			Hellobacteriaceae	Staphylococcus	0.18%
			Leptospiraceae	Chlamydia	0.16%
			Micrococccaceae	Pyrococcus	0.16%
			Nautiliaceae	Sodalis	0.16%
			Nitrosopumilaceae	Bdellovibrio	0.14%
			Nitrosospiraceae	Desulfovibrio	0.14%
			Rhodobacteraceae	Ehrlichia	0.14%
			Sphingomonadaceae	Jarthinobacterium	0.14%
				Metallosphaera	0.14%
				Thermoanaerobacter	0.14%
				Candidatus Koribacter	0.12%
				Geobacillus	0.12%
				Gramella	0.12%
				Nostoc	0.12%
				Pediococcus	0.12%
				Propionibacterium	0.12%
				Proteus	0.12%
				Syntrophomonas	0.12%
				Wolbachia	0.12%
				Candidatus Blochmannia	0.10%
				Chloroherpeton	0.10%
				Colwellia	0.10%
				Fingoldia	0.10%
				Hahella	0.10%
				Herpetosiphon	0.10%
				Methanosphaerula	0.10%
				Natronaerobius	0.10%
				Salmonella	0.10%
				Sulfolobus	0.10%
				Thiomicrospira	0.10%
				Allivibrio	0.08%
				Parabacteroides	0.08%
				Anoxybacillus	0.06%
				Dechloromonas	0.06%
				Enterococcus	0.06%
				Francisella	0.06%
				Methanocorpusculum	0.06%
				Methanopyrus	0.06%
				Methanospirillum	0.06%
				Methanothermobacter	0.06%
				Oenococcus	0.06%
				Pasteurella	0.06%
				Photorhabdus	0.06%
				Aeropyrum	0.04%
				Borrelia	0.04%
				Caldicellulosiruptor	0.04%
				Caldivirga	0.04%
				Candidatus Protochlamydia	0.04%
				Haemophilus	0.04%
				Klebsiella	0.04%
				Methanococcus	0.04%
				Methylophilum	0.04%
				Pelotomaculum	0.04%
				Thermofilum	0.04%
				Actinobacillus	0.02%
				Anabaena	0.02%
				Campylobacter	0.02%
				Chromohalobacter	0.02%
				Cyanothece	0.02%
				Dehalococcoides	0.02%
				Deinococcus	0.02%
				Desulfobacterium	0.02%
				Desulfurococcus	0.02%
				Dinoroseobacter	0.02%
				Enterobacter	0.02%
				Halorcula	0.02%
				Hellobacterium	0.02%
				Herminiimonas	0.02%
				Leptospira	0.02%
				Moorella	0.02%
				Nautilia	0.02%
				Nitratiruptor	0.02%
				Nitrosopumilus	0.02%
				Novosphingobium	0.02%
				Renibacterium	0.02%
				Thermodesulfovibrio	0.02%
				Thermosipho	0.02%
				Thermosynechococcus	0.02%
				Trichodesmium	0.02%

表六、*Prevotella melaninogenica* 在各分類階層的分類情形

Phylum	Percentag	Class	Percentag	Order	Percentag	Family	Percentag	Genus	Percentag
Fimicutes	43.08%	Bacilli	27.40%	Lactobacillales	18.84%	Streptococcaceae	17.36%	Streptococcus	17.34%
Proteobacteria	27.32%	Clostridia	15.68%	Rickettsiales	10.78%	Anaplasmataceae	8.80%	Bacteroides	7.26%
Bacteroidetes	10.70%	Gammaproteobacteria	12.92%	Bacteroidales	9.72%	Bacteroidaceae	7.26%	Vibrio	6.40%
Euryarchaeota	7.98%	Alphaproteobacteria	12.42%	Clostridiales	9.70%	Vibrionaceae	7.16%	Coprothermobacter	4.94%
Thermotogae	2.68%	Bacteroidia	9.72%	Bacillales	8.56%	Clostridiaceae	5.60%	Neorickettsia	4.64%
Chlamydiae	2.04%	Methanomicrobia	4.54%	Vibrionales	7.16%	Thermodesulfobiaceae	4.94%	Methanococcoides	4.18%
Crenarchaeota	1.96%	Thermotogae	2.68%	Thermoanaerobacterales	5.86%	Methanosarcinaceae	4.48%	Clostridium	4.08%
Chlorobi	1.34%	Thermoprotei	1.96%	Methanosarcinales	4.50%	Staphylococcaceae	4.12%	Eubacterium	3.54%
Cyanobacteria	1.22%	Thermoplasmata	1.40%	Alteromonadales	2.98%	Bacillaceae	3.88%	Anaplasma	3.38%
Spirochaetes	0.68%	Chlorobia	1.34%	Thermotogales	2.68%	Eubacteriaceae	3.54%	Macrococcus	3.08%
Actinobacteria	0.44%	Deltaproteobacteria	0.96%	Chlamydiales	2.04%	Thermotogaceae	2.68%	Shewanella	2.10%
Aquificae	0.22%	Halobacteria	0.92%	Sulfobacterales	1.92%	Shewanellaceae	2.10%	Bacillus	2.04%
Tenericutes	0.12%	Epsilonproteobacteria	0.78%	Rhizobiales	1.64%	Rickettsiaceae	1.98%	Sulfolobus	1.74%
Dictyoglomi	0.08%	Thermococci	0.56%	Thermoplasmatales	1.40%	Sulfobacterae	1.92%	Bartonella	1.64%
Thaumarchaeota	0.08%	Archaeoglobi	0.50%	Chlorobiales	1.34%	Chlamydiaceae	1.72%	Wolbachia	1.54%
Nitrospirae	0.04%	Actinobacteria	0.44%	Halobacteriales	0.92%	Bartonellaceae	1.64%	Thermoplasma	1.40%
Verrucomicrobia	0.02%	Betaproteobacteria	0.24%	Spirochaetales	0.68%	Thermoplasmataceae	1.40%	Lysinibacillus	1.28%
		Aquificae	0.22%	Thermococcales	0.56%	Chlorobiaceae	1.34%	Porphyromonas	1.18%
		Mollicutes	0.12%	Prochlorales	0.54%	Porphyromonadaceae	1.32%	Candidatus Azobacteroides	1.14%
		Dictyoglomia	0.08%	Archaeoglobales	0.50%	Halobacteriaceae	0.92%	Staphylococcus	1.04%
		Methanococci	0.06%	Legionellales	0.48%	Lactobacillaceae	0.82%	Fervidobacterium	1.02%
		Nitrospira	0.04%	Oceanospirillales	0.46%	Spirochaetaceae	0.62%	Alkaliphilus	0.94%
		Gloeobacteria	0.02%	Desulfovibrionales	0.44%	Thermoanaerobacteraceae	0.58%	Haloquadratum	0.92%
				Enterobacteriales	0.44%	Thermococcaceae	0.56%	Chlamydomonas	0.88%
				Psuedomonadales	0.44%	Prochlorococcaceae	0.54%	Pelodictyon	0.88%
				Desulfobacteriales	0.40%	Leuconostocaceae	0.52%	Chlamydia	0.84%
				Thiotrichales	0.40%	Archaeoglobaceae	0.50%	Lactobacillus	0.80%
				Actinomycetales	0.36%	Legionellaceae	0.46%	Ehrlichia	0.78%
				Flavobacteriales	0.36%	Oceanospirillaceae	0.46%	Kosmotoga	0.60%
				Nostocales	0.36%	Desulfovibrionaceae	0.44%	Thermoanaerobacter	0.58%
				Aquificales	0.22%	Enterobacteriaceae	0.44%	Oceanobacillus	0.56%
				Pasteurellales	0.16%	Moraxellaceae	0.44%	Photobacterium	0.56%
				Chroococcales	0.12%	Desulfobulbaceae	0.40%	Prochlorococcus	0.54%
				Desulfuromonadales	0.12%	Flavobacteriaceae	0.36%	Candidatus Amoebophilus	0.52%
				Natranaerobiales	0.12%	Francisellaceae	0.36%	Leuconostoc	0.52%
				Nitrosomonadales	0.12%	Nostocaceae	0.36%	Sulfurovum	0.52%
				Acholeplasmatales	0.10%	Paenibacillaceae	0.36%	Archaeoglobus	0.50%
				Methylophilales	0.10%	Thermoanaerobacterales	0.34%	Legionella	0.46%
				Sphingobacteriales	0.10%	Alteromonadaceae	0.32%	Marinomonas	0.46%
				Bifidobacteriales	0.08%	Clostridiales	0.32%	Petrogla	0.46%
				Dictyoglomales	0.08%	Colwelliaceae	0.30%	Thermotoga	0.46%
				Nitrospumillales	0.08%	Peptococcaceae	0.22%	Lawsonia	0.42%
				Xanthomonadales	0.08%	Corynebacteriaceae	0.18%	Desulfotalea	0.40%
				Campylobacteriales	0.06%	Hydrogenothermaceae	0.18%	Acinetobacter	0.38%
				Methanococcales	0.06%	Pseudoalteromonadaceae	0.18%	Rickettsia	0.38%
				Nautilliales	0.06%	Parachlamydiaceae	0.16%	Brevibacillus	0.36%
				Oscillatoriales	0.06%	Pasteurellaceae	0.16%	Flavobacterium	0.36%
				Methanomicrobiales	0.04%	Enterococcaceae	0.14%	Francisella	0.36%
				Nitrospirales	0.04%	Geobacteraceae	0.12%	Caldicellulosiruptor	0.34%
				Burkholderiales	0.02%	Natranaerobiaceae	0.12%	Alteromonas	0.32%
				Chromatiales	0.02%	Nitrosomonadaceae	0.12%	Finexoldia	0.32%
				Desulfurococcales	0.02%	Acholeplasmataceae	0.10%	Treponema	0.32%
				Gloeobacteriales	0.02%	Methylophilaceae	0.10%	Borrelia	0.30%
				Methylacidiphilales	0.02%	Bifidobacteriaceae	0.08%	Chlorobium	0.30%
				Mycoplasmatales	0.02%	Dictyoglomaceae	0.08%	Colwellia	0.30%
				Thermoproteales	0.02%	Listeriaceae	0.08%	Methanosarcina	0.30%
						Nitrospumillaceae	0.08%	Pyrococcus	0.30%
						Xanthomonadaceae	0.08%	Nostoc	0.26%
						Helicobacteraceae	0.06%	Thermococcus	0.26%
						Leptospiraceae	0.06%	Proteus	0.24%
						Methanococcaceae	0.06%	Allivibrio	0.20%
						Nautillaceae	0.06%	Corynebacterium	0.18%
						Psychromonadaceae	0.06%	Metallosphaera	0.18%
						Aquificaceae	0.04%	Pseudoalteromonas	0.18%
						Methanocorpusculaceae	0.04%	Candidatus Protochlamydia	0.16%
						Nitrospiraceae	0.04%	Desulfotomaculum	0.16%
						Piscirickettsiaceae	0.04%	Prosthecochloris	0.16%

表六、*Prevotella melaninogenica* 在各分類階層的分類情形(續)

Phylum	Percentag Class	Percentag Order	Percentag Family	Percentag Genus	Percentag	
			Burkholderiaceae	0.02%	Sulfurihydrogenibium	0.16%
			Chromatiaceae	0.02%	Tropheryma	0.16%
			Coxiellaceae	0.02%	Enterococcus	0.14%
			Desulfurococcaceae	0.02%	Nitratiruptor	0.14%
			Hellobacteriaceae	0.02%	Parabacteroides	0.14%
			Idiomarinaceae	0.02%	Thermosiphon	0.14%
			Methanosphaeraceae	0.02%	Acaryochloris	0.12%
			Methylacidiphilaceae	0.02%	Exiguobacterium	0.12%
			Mycoplasmataceae	0.02%	Geobacter	0.12%
			Propionibacteriaceae	0.02%	Natronaerobius	0.12%
			Thermoproteaceae	0.02%	Acholeplasma	0.10%
					Anabaena	0.10%
					Cytophaga	0.10%
					Haemophilus	0.10%
					Methylobacillus	0.10%
					Bifidobacterium	0.08%
					Candidatus Blochmannia	0.08%
					Dictyoglomus	0.08%
					Listeria	0.08%
					Nitrosomonas	0.08%
					Nitrosopumilus	0.08%
					Synechococcus	0.08%
					Xylella	0.08%
					Desulfitobacterium	0.06%
					Leptospira	0.06%
					Methanococcus	0.06%
					Nautilia	0.06%
					Orientia	0.06%
					Pasteurella	0.06%
					Pectobacterium	0.06%
					Psychrobacter	0.06%
					Psychromonas	0.06%
					Trichodesmium	0.06%
					Yersinia	0.06%
					Cyanothece	0.04%
					Hydrogenobaculum	0.04%
					Methanocorpusculum	0.04%
					Nitrospira	0.04%
					Sulfurimonas	0.04%
					Thermodesulfobacterium	0.04%
					Thiomicrospira	0.04%
					Coxiella	0.02%
					Desulfobacterium	0.02%
					Desulfurococcus	0.02%
					Gloeobacter	0.02%
					Helicobacter	0.02%
					Hellobacterium	0.02%
					Idiomarina	0.02%
					Lactococcus	0.02%
					Methanosarcina	0.02%
					Methylacidiphilum	0.02%
					Mycoplasma	0.02%
					Nitrosococcus	0.02%
					Pediococcus	0.02%
					Persephonella	0.02%
					Polynucleobacter	0.02%
					Propionibacterium	0.02%
					Pyrobaculum	0.02%

表七、Lactococcus garvieae Lg2 在各分類階層的分類情形

Phylum	Percentage Class	Percentage Order	Percentage Family	Percentage Genus	Percentage				
Firmicutes	59.94%	Bacilli	51.52%	Lactobacillales	43.74%	Streptococcaceae	36.52%	Streptococcus	28.84%
Proteobacteria	21.46%	Gamma proteobacteria	9.88%	Bacillales	7.78%	Bartonellaceae	6.42%	Lactococcus	7.68%
Bacteroidetes	6.66%	Alpha proteobacteria	9.44%	Rhizobiales	7.10%	Bacillaceae	3.32%	Bartonella	6.42%
Chlamydiae	3.30%	Clostridia	7.64%	Clostridiales	4.26%	Lactobacillaceae	2.92%	Lactobacillus	2.76%
Thermotogae	2.54%	Chlamydia	3.30%	Chlamydiales	3.30%	Moraxellaceae	2.56%	Bacillus	2.74%
Cyanobacteria	1.40%	Flavobacteria	2.84%	Thermoanaerobacterales	3.16%	Thermotogaceae	2.54%	Acinetobacter	2.50%
Euryarchaeota	1.20%	Thermotogae	2.54%	Flavobacteriales	2.84%	Clostridiaceae	2.46%	Enterococcus	2.10%
Spirochaetes	0.82%	Bacteroidia	2.14%	Pseudomonadales	2.56%	Flavobacteriaceae	2.32%	Listeria	1.92%
Tenericutes	0.46%	Cytophagia	1.30%	Thermotogales	2.54%	Enterococcaceae	2.16%	Clostridium	1.82%
Actinobacteria	0.38%	Epsilonproteobacteria	1.02%	Bacteroidales	2.14%	Staphylococcaceae	1.94%	Staphylococcus	1.64%
Deferribacteres	0.38%	Beta proteobacteria	0.84%	Rickettsiales	1.50%	Listeriaceae	1.92%	Legionella	1.32%
Fusobacteria	0.28%	Spirochaetia	0.82%	Legionellales	1.48%	Chlamydiaceae	1.34%	Leucostoc	1.14%
Elusimicrobia	0.20%	Erysipelotrichi	0.74%	Cytophagales	1.30%	Leucostocaceae	1.34%	Thermotoga	1.08%
Aquificae	0.16%	Mollicutes	0.46%	Vibrionales	1.20%	Legionellaceae	1.32%	Caldicellulosiruptor	0.88%
Crenarchaeota	0.16%	Methanomicrobia	0.40%	Pasteurellales	1.18%	Anaplasmataceae	1.24%	Chlamydia	0.80%
Chlorobi	0.14%	Actinobacteria	0.38%	Enterobacteriales	1.06%	Vibrionaceae	1.20%	Erysipelothrix	0.74%
Synergistetes	0.12%	Deferribacteres	0.38%	Oceanospirillales	0.96%	Pasteurellaceae	1.18%	Prochlorococcus	0.74%
Nitrospirae	0.10%	Sphingobacteriia	0.34%	Alteromonadales	0.84%	Thermoanaerobacterales	1.18%	Candidatus Protoclhamydia	0.72%
Thermodesulfobacteria	0.10%	Fusobacteriia	0.28%	Spirochaetales	0.82%	Thermoanaerobacteraceae	1.12%	Simkania	0.70%
Fibrobacteres	0.08%	Thermococci	0.28%	Erysipelotrichales	0.74%	Enterobacteriaceae	1.06%	Candidatus Liberibacter	0.68%
Chloroflexi	0.02%	Methanobacteria	0.24%	Prochlorales	0.74%	Parachlamydiaceae	1.04%	Ferribacterium	0.68%
Dictyoglomi	0.02%	Deltaproteobacteria	0.22%	Campylobacteriales	0.72%	Thermodesulfobiaceae	0.86%	Vibrio	0.66%
		Methanococci	0.22%	Rhodobacteriales	0.62%	Erysipelotrichaceae	0.74%	Wolbachia	0.66%
		Elusimicrobia	0.20%	Burkholderiales	0.50%	Prochlorococcaceae	0.74%	Hirschia	0.62%
		Aquificae	0.16%	Chroococcales	0.42%	Simkaniaceae	0.70%	Haemophilus	0.58%
		Themoprotei	0.16%	Methanosarcinales	0.40%	Rhizobiaceae	0.68%	Mairiviga	0.58%
		Chlorobia	0.14%	Thiotrichales	0.40%	Spirochaetaceae	0.68%	Alkaliphilus	0.56%
		Synergistia	0.12%	Deferribacterales	0.38%	Hyphomonadaceae	0.62%	Cyclobacterium	0.56%
		Nitrospira	0.10%	Mycoplasmatales	0.36%	Flammovirgaceae	0.58%	Chlamydomphila	0.54%
		Thermodesulfobacteria	0.10%	Sphingobacteriales	0.34%	Cyclobacteriaceae	0.56%	Carnobacterium	0.50%
		Fibrobacteriia	0.08%	Fusobacteriales	0.28%	Helicobacteraceae	0.52%	Fluviolca	0.50%
		Negativicutes	0.04%	Thermococcales	0.28%	Camobacteriaceae	0.50%	Borrelia	0.48%
		Archaeoglobi	0.02%	Methanobacteriales	0.24%	Cryomorphaceae	0.50%	Kangielia	0.48%
		Dehalococcoidetes	0.02%	Coriobacteriales	0.22%	Porphyromonadaceae	0.50%	Marinomonas	0.48%
		Dictyoglomia	0.02%	Methanococcales	0.22%	Alcanivoracaceae	0.48%	Neorickettsia	0.48%
				Nitrosomonadales	0.20%	Oceanospirillaceae	0.48%	Paludibacter	0.48%
				Natranerobiales	0.18%	Paenibacillaceae	0.44%	Sulfurimonas	0.48%
				Sphingomonadales	0.18%	Methanosarcinaceae	0.40%	Coprotethermobacter	0.46%
				Aquificales	0.16%	Peptococcaceae	0.40%	Thermoanaerobacter	0.46%
				Nostocales	0.16%	Bacteroidaceae	0.38%	Candidatus Hamiltonella	0.40%
				Sulfobacterales	0.16%	Deferribacteraceae	0.38%	Muricauda	0.40%
				Chlorobiales	0.14%	Mycoplasmataceae	0.36%	Tepidanaerobacter	0.40%
				Methylophilales	0.14%	Lachnospiraceae	0.34%	Thermodesulfobium	0.40%
				Nautiliales	0.14%	Sphingobacteriaceae	0.34%	Bacteroides	0.38%
				Bifidobacteriales	0.12%	Burkholderiaceae	0.32%	Flavobacterium	0.38%
				Chromatiales	0.12%	Aerococcaceae	0.30%	Lysinibacillus	0.36%
				Synergistales	0.12%	Clostridiales	0.30%	Paenibacillus	0.36%
				Acholeplasmatales	0.10%	Piscirickettsiaceae	0.30%	Petrogoga	0.36%
				Desulfobiontales	0.10%	Shewanellaceae	0.30%	Histophilus	0.34%
				Desulfurellales	0.10%	Thermococcaceae	0.28%	Mycoplasma	0.34%
				Nitrospirales	0.10%	Coriobacteriaceae	0.22%	Pedobacter	0.34%
				Thermodesulfobacteriales	0.10%	Waddliaceae	0.22%	Wanibacter	0.32%
				Fibrobacteriales	0.08%	Campylobacteraceae	0.20%	Parachlamydia	0.32%
				Oscillatoriales	0.08%	Colwelliaceae	0.20%	Polynucleobacter	0.32%
				Actinomycetales	0.04%	Methanococcaceae	0.20%	Aerococcus	0.30%
				Elusimicrobiales	0.04%	Nitrosomonadaceae	0.20%	Macrococcus	0.30%
				Halanaerobiales	0.04%	Rickettsiaceae	0.20%	Shewanella	0.30%
				Selenomonadales	0.04%	Alcaligenaceae	0.18%	Thermoanaerobacterium	0.30%
				Xanthomonadales	0.04%	Leptotrichiaceae	0.18%	Thiomicrospira	0.30%
				Archaeoglobales	0.02%	Methanobacteriaceae	0.18%	Methanosarcina	0.28%
				Desulfobacteriales	0.02%	Natranerobiaceae	0.18%	Aliivibrio	0.26%
				Dictyoglomales	0.02%	Sphingomonadaceae	0.18%	Thermococcus	0.26%
						Alteromonadaceae	0.16%	Caldanaerobacter	0.24%
						Coxiellaceae	0.16%	Finogdia	0.24%
						Cytophagaceae	0.16%	Themosipho	0.24%
						Nostocaceae	0.16%	Pasteurella	0.22%
						Peptostreptococcaceae	0.16%	Waddlia	0.22%
						Sulfobacterae	0.16%	Weeksella	0.22%
						Chlorobiaceae	0.14%	Atopobium	0.20%
						Methylophilaceae	0.14%	Colwellia	0.20%
						Nautiliaceae	0.14%	Flexistipes	0.20%
						Prevotellaceae	0.14%	Nitrosomonas	0.20%
						Bifidobacteriaceae	0.12%	Treponema	0.20%
						Chromatiaceae	0.12%	Cellulosilyticum	0.18%
						Synergistaceae	0.12%	Desulfobacterium	0.18%

表七、*Lactococcus garvieae* Lg2 在各分類階層的分類情形(續)

Phylum	Percentage Class	Percentage Order	Percentage Family	Percentage Genus	Percentage	
			Acholeplasmataceae	0.10%	Kosmotoga	0.18%
			Desulfobionaceae	0.10%	Methanococcus	0.18%
			Desulfurellaceae	0.10%	Natranaerobius	0.18%
			Francisellaceae	0.10%	Rickettsia	0.18%
			Fusobacteriaceae	0.10%	Taylorella	0.18%
			Leptospiraceae	0.10%	Xenorhabdus	0.18%
			Nitrospiraceae	0.10%	Zunongwangia	0.18%
			Thermodesulfobacteriaceae	0.10%	Zymomonas	0.18%
			Desulfobacteriaceae	0.08%	Butyrivibrio	0.16%
			Fibrobacteraceae	0.08%	Cellulophaga	0.16%
			Hydrogenothermaceae	0.08%	Coxiella	0.16%
			Psychromonadaceae	0.08%	Exiguobacterium	0.16%
			Methanothermaceae	0.06%	Leptotrichia	0.16%
			Pseudoalteromonadaceae	0.06%	Pediococcus	0.16%
			Elusimicrobiaceae	0.04%	Photorhabdus	0.16%
			Idiomarinaceae	0.04%	Alteromonas	0.14%
			Syntrophomonadaceae	0.04%	Cyanotheca	0.14%
			Veillonellaceae	0.04%	Desulfotomaculum	0.14%
			Xanthomonadaceae	0.04%	Listonella	0.14%
			Archaeoglobaceae	0.02%	Methylotenera	0.14%
			Blattabacteriaceae	0.02%	Nautilia	0.14%
			Corynebacteriaceae	0.02%	Oenococcus	0.14%
			Desulfobulbaceae	0.02%	Photobacterium	0.14%
			Dictyoglomaceae	0.02%	Prevotella	0.14%
			Eubacteriaceae	0.02%	Sulfolobus	0.14%
			Halanaerobiaceae	0.02%	Zobellia	0.14%
			Halobacteroidaceae	0.02%	Aminobacterium	0.12%
			Methanocaldococcaceae	0.02%	Candidatus Azobacteroides	0.12%
					Capnocytophaga	0.12%
					Gardnerella	0.12%
					Nitratiruptor	0.12%
					Nitrosococcus	0.12%
					Sulfurospirillum	0.12%
					Candidatus Blochmannia	0.10%
					Cytophaga	0.10%
					Francisella	0.10%
					Geobacillus	0.10%
					Hippea	0.10%
					Ilyobacter	0.10%
					Leptospira	0.10%
					Methanobacterium	0.10%
					Thermodesulfatator	0.10%
					Thermodesulfovibrio	0.10%
					Brevibacillus	0.08%
					Buchnera	0.08%
					Campylobacter	0.08%
					Candidatus Arthromitus	0.08%
					Croceibacter	0.08%
					Denitrovibrio	0.08%
					Desulfobacterium	0.08%
					Fibrobacter	0.08%
					Gramella	0.08%
					Krokinobacter	0.08%
					Methanobrevibacter	0.08%
					Methanococcoides	0.08%
					Nostoc	0.08%
					Oceanobacillus	0.08%
					Psychromonas	0.08%
					Riemerella	0.08%
					Sulfurihydrogenibium	0.08%
					Syntrophobotulus	0.08%
					Thermobaculum	0.08%
					Trichodesmium	0.08%
					Anaerococcus	0.06%
					Calditerrivibrio	0.06%
					Candidatus Midichloria	0.06%
					Candidatus Phytoplasma	0.06%
					Chlorobium	0.06%
					Chloroherpeton	0.06%
					Desulfovibrio	0.06%
					Ehrlichia	0.06%
					Magnetococcus	0.06%
					Melissococcus	0.06%
					Methanothermus	0.06%
					Microcystis	0.06%
					Proteus	0.06%

表七、*Lactococcus garvieae* Lg2 在各分類階層的分類情形(續)

Phylum	Percentage Class	Percentage Order	Percentage Family	Percentage Genus	Percentage
				Pseudoalteromonas	0.06%
				Psychrobacter	0.06%
				Weissella	0.06%
				Yersinia	0.06%
				Acholeplasma	0.04%
				Aciduliprofundum	0.04%
				Anabaena	0.04%
				Anaplasma	0.04%
				Anoxybacillus	0.04%
				Candidatus Amoebophilus	0.04%
				Candidatus Pelagibacter	0.04%
				Defimbacter	0.04%
				Elusimicrobium	0.04%
				Idiomarina	0.04%
				Lawsonia	0.04%
				Leadbetterella	0.04%
				Sphaerochaeta	0.04%
				Sulfurivum	0.04%
				Sulfurovum	0.04%
				Syntrophothermus	0.04%
				Themosynechococcus	0.04%
				Trichormus	0.04%
				Veillonella	0.04%
				Xylella	0.04%
				Acetohalobium	0.02%
				Actinobacillus	0.02%
				Archaeoglobus	0.02%
				Blattabacterium	0.02%
				Candidatus Riesia	0.02%
				Carboxydotherrus	0.02%
				Corynebacterium	0.02%
				Cryptobacterium	0.02%
				Dehalococcoides	0.02%
				Desulfotalea	0.02%
				Dictyoglomus	0.02%
				Eubacterium	0.02%
				Gallibacterium	0.02%
				Glaciecola	0.02%
				Halanaerobium	0.02%
				Lacinutrix	0.02%
				Metallosphaera	0.02%
				Methanocaldococcus	0.02%
				Methanohalobium	0.02%
				Methanosalsum	0.02%
				Methanothermococcus	0.02%
				Orientia	0.02%
				Pelodictyon	0.02%
				Porphyromonas	0.02%
				Pyrococcus	0.02%
				Runella	0.02%
				Sebalidella	0.02%
				Tropheryma	0.02%
				Ureaplasma	0.02%

表八、不同操作方法是否影響無質體序列資料庫的分類準確度

Comparing supervised and semi-supervised difference							
Taxonomy	Superkingdom	Phylum	Class	Order	Family	Genus	Species
t-test	NA	-	-	-	NA	NA	NA

比較在 supervised 和 semi-supervised 的操作方式下，未加入質體序列的原核生物序列特徵資料庫在各分類層級中分類原核生物序列的準確度上，是否有所差異。NA 表示無法比較，“-”表示無顯著差異，“+”表示有顯著差異。

表九、不同操作方法是是否影響真核資料庫的分類準確度

Comparing supervised and semi-supervised difference							
Taxonomy	Kingdom	Phylum	Class	Order	Family	Genus	Species
t-test	-	-	-	-	-	-	-

比較在 supervised 和 semi-supervised 的方式下，真核生物序列特徵資料庫在各階層中分類真核生物序列的準確度上，是否有所差異。NA 表示無法比較，“-“表示無顯著差異，“+“表示有顯著差異。

表十、不同操作方法是是否影響原核資料庫分類混合序列的結果

Comparing supervised and semi-supervised difference							
Taxonomy	Kingdom	Phylum	Class	Order	Family	Genus	Species
t-test	-	++	++	++	++	++	++

比較在 supervised 和 semi-supervised 的方式下，真核序列特徵資料庫分類混合序列的結果，在各分類階層中是否有所差異。NA 表示無法比較，“-“表示無顯著差異，“+“表示有顯著差異。

表十一、比較混合與真核序列的分類結果

Comparing the differences between mix and eukaryotic sequences accuracy rate in two models								
Taxonomy		Kingdom	Phylum	Class	Order	Family	Genus	Species
t-test	supervised	NA	NA	-	-	-	-	-
	semi-supervised	-	++	++	++	++	++	++

對真核生物序列特徵資料庫分類混合序列的結果與單獨只分類真核生物序列的結果，兩者進行比較。NA 表示無法比較，“-“表示無顯著差異，“+“表示有顯著差異。

表十二、不同操作方法是否影響原核資料庫分類混合序列的結果

Comparing supervised and semi-supervised difference							
Taxonomy	Superkingdom	Phylum	Class	Order	Family	Genus	Species
t-test	-	-	-	-	++	-	++

比較在 supervised 和 semi-supervised 的方式下，原核序列特徵資料庫分類混合序列的結果，在各分類階層中是否有所差異。NA 表示無法比較，“-”表示無顯著差異，“+”表示有顯著差異。

表十三、比較混合與原核序列的分類結果

Comparing the differences between mix and prokaryotic sequences accuracy rate in two models								
Taxonomy		Superkingdom	Phylum	Class	Order	Family	Genus	Species
t-test	supervised	-	-	-	-	+	+	++
	semi-supervised	-	-	-	-	+	+	++

對未加質體的原核生物序列特徵資料庫分類混合序列的結果與單獨只分類原核生物序列的結果，兩者進行比較。NA表示無法比較，“-”表示無顯著差異，“+”表示有顯著差異。

表十四、不同操作方法是否影響不包含於資料庫的種序列分析

Comparing supervised and semi-supervised difference						
Taxonomy	Superkingdom	Phylum	Class	Order	Family	Genus
t-test	-	+	++	++	-	-

比較在 supervised 和 semi-supervised 的方式下，未加入質體序列的原核生物序列特徵資料庫在各分類層級中分類非資料庫的原核生物序列的準確度上，是否有所差異。NA 表示無法比較，” - “表示無顯著差異， “+ “表示有顯著差異。

表十五、不同操作方法是否影響不包含於資料庫的屬序列分析

Comparing supervised and semi-supervised difference					
Taxonomy	Superkingdom	Phylum	Class	Order	Family
t-test	+	++	++	++	++

比較在 supervised 和 semi-supervised 的方式下，未加入質體序列的原核生物序列特徵資料庫在各分類層級中分類非資料庫屬的原核生物序列的準確度上，是否有所差異。NA 表示無法比較，” - “表示無顯著差異， “+ “表示有顯著差異。

個人資料

中文姓名：吳怡靜

英文姓名：I-Ching Wu

出生日期：1988 年 4 月 23 日

戶籍地址：苗栗縣三義鄉廣盛村廣聲新城 39 號

聯絡電話：037-872650

電子信箱：blue90211@gmail.com

學歷：

東海大學生命科學系學士

東海大學生命科學系碩士