

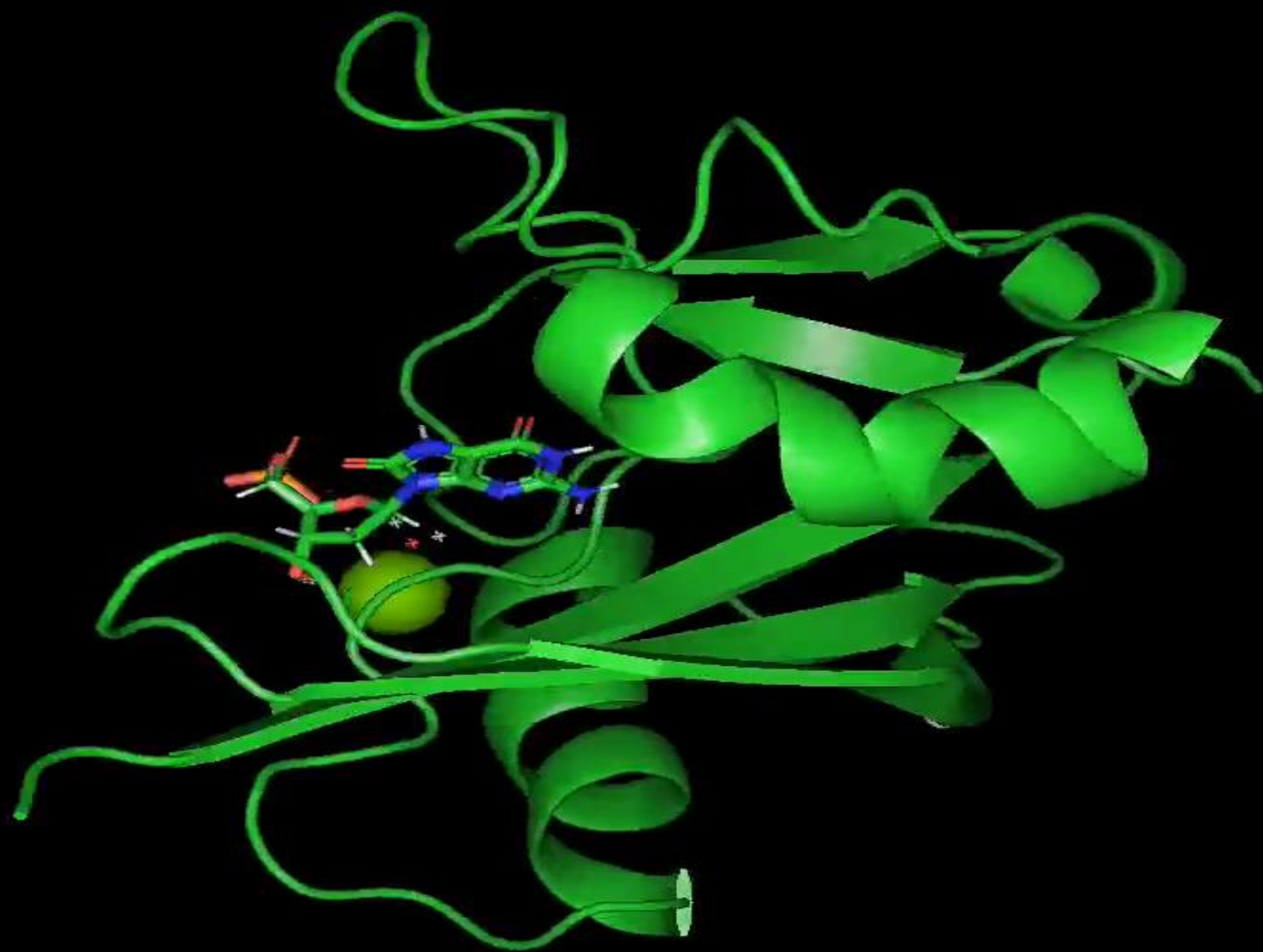
Εύρηκα

NUDIX水解酶在 大腸桿菌中的演化

連啟竣
鄭貽生 教授
魏宏仁 老師









Εύρηκα

定義一

Superfamily

蛋白質超家族

具有共同祖先的蛋白質群

動機:

Εύρηκα

演化是否存在一個規律

功能、結構、序列...





Εύρηκα

第一步：定義問題方向



Εύρηκα

大腸桿菌的NUDIX
Hydrolase的演化是
否存在一個規律



Eύρηκα

定義二 Motif 結構域

蛋白質中，具有共同胺基酸
序列之區域



定義三

Εύρηκα

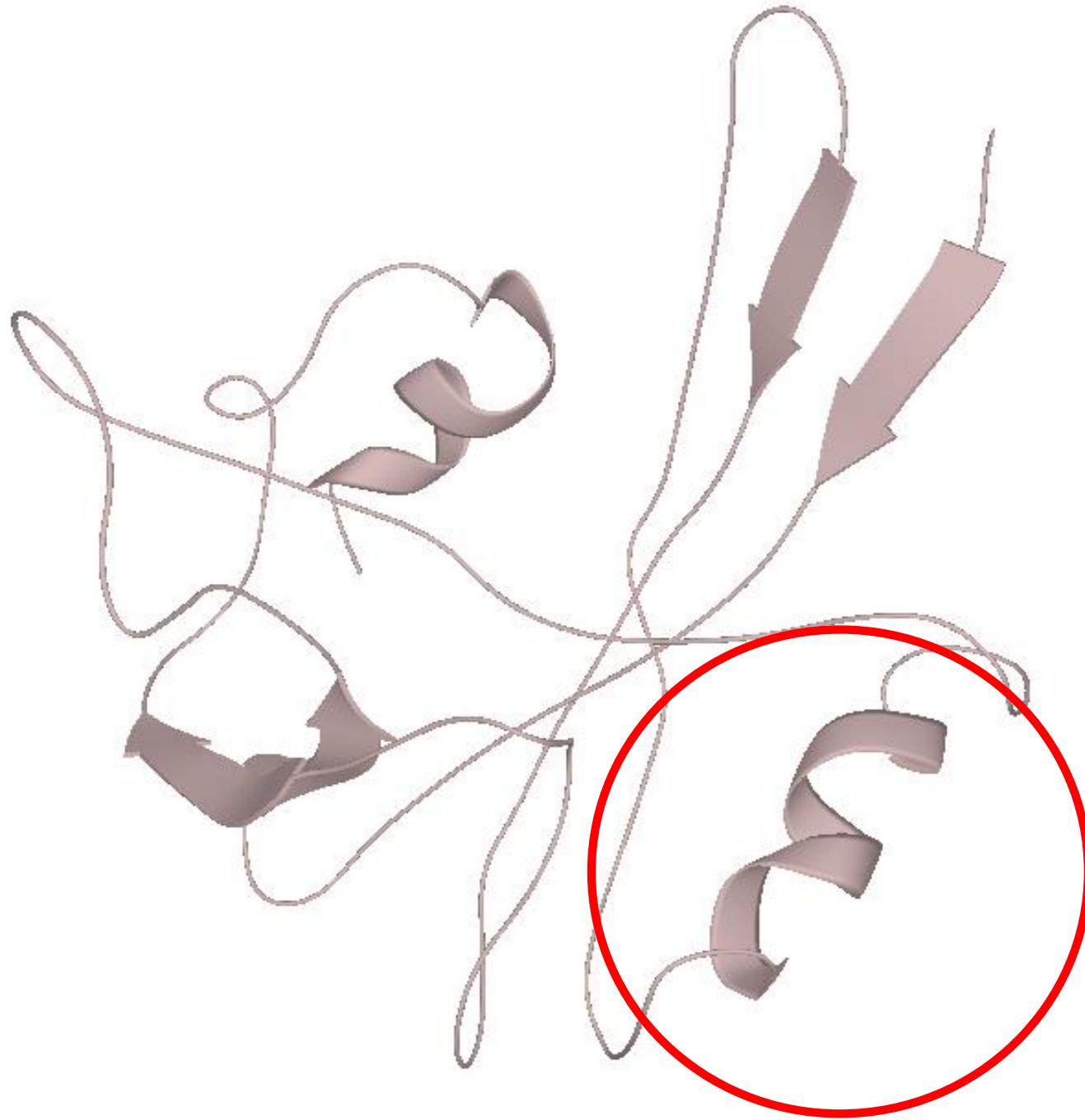
NUDIX hydrolase

NUDIX 水解酶

具有特定MOTIF的SUPERFAMILY，
並水解核苷酸衍生物



MuTT



$\gamma K\alpha$

β

Εύρηκα

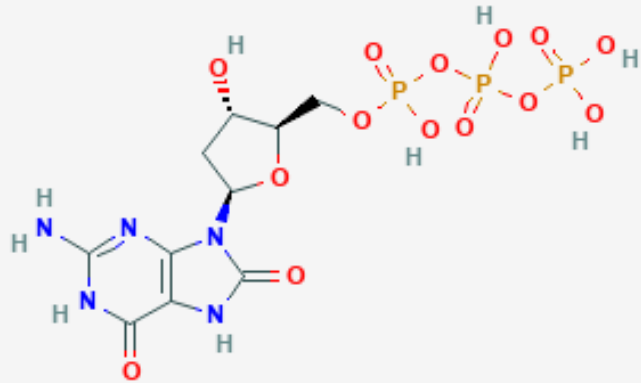
G ⊙ ⊙ ⊙ ⊙ ⊙ E ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ REU ⊙ EE ⊙ GU

將成為NUDIX BOX 所
在的位置之分辨依據

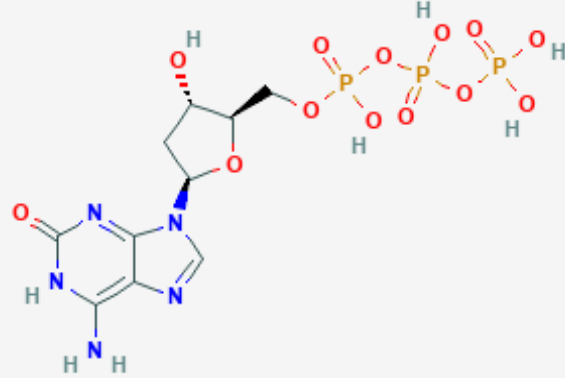


Εύρηκα

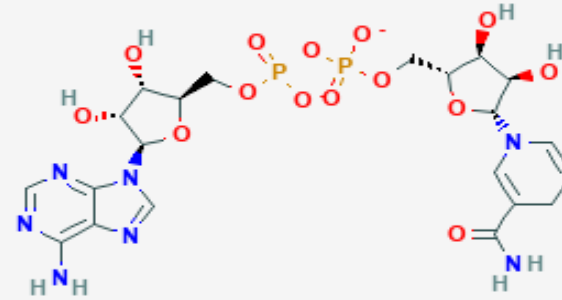
文物继续度高



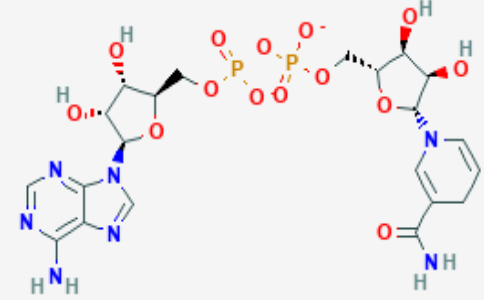
8-oH-dGTP



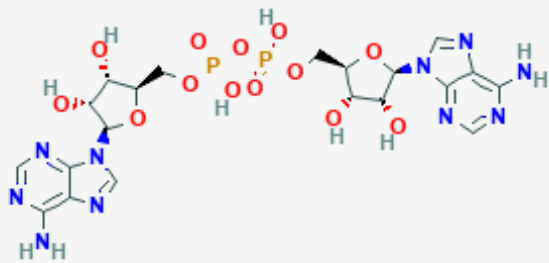
2-OH-dATP



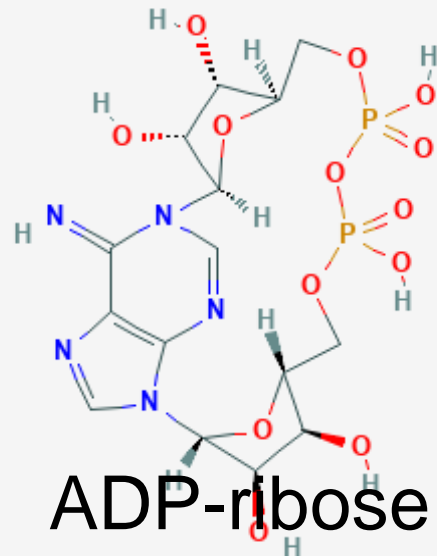
NADH



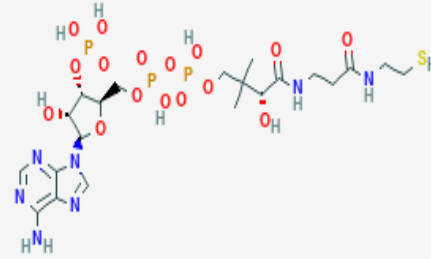
GDP-mannose



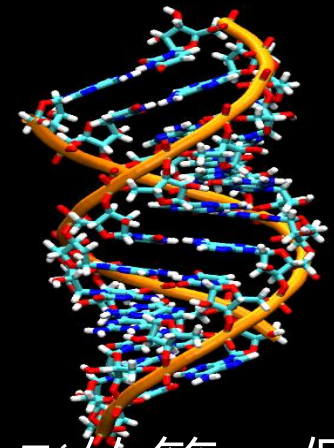
Ap2A



ADP-ribose



輔酶A



RNA 5' 的第一個鹼基

Εύρηκα

是否存在一個分類的方法？

是否可以圖像化？

並藉由圖像尋找演化的趨勢

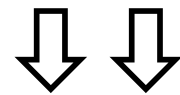


對於每一個物種，
同時具有不同功能的NUDIX

每個NUDIX是為
細胞功能而演化



類似的序列意味
類似的功能



透過不同功能的蛋白質的序列，
尋找其演化的趨勢



Εύρηκα

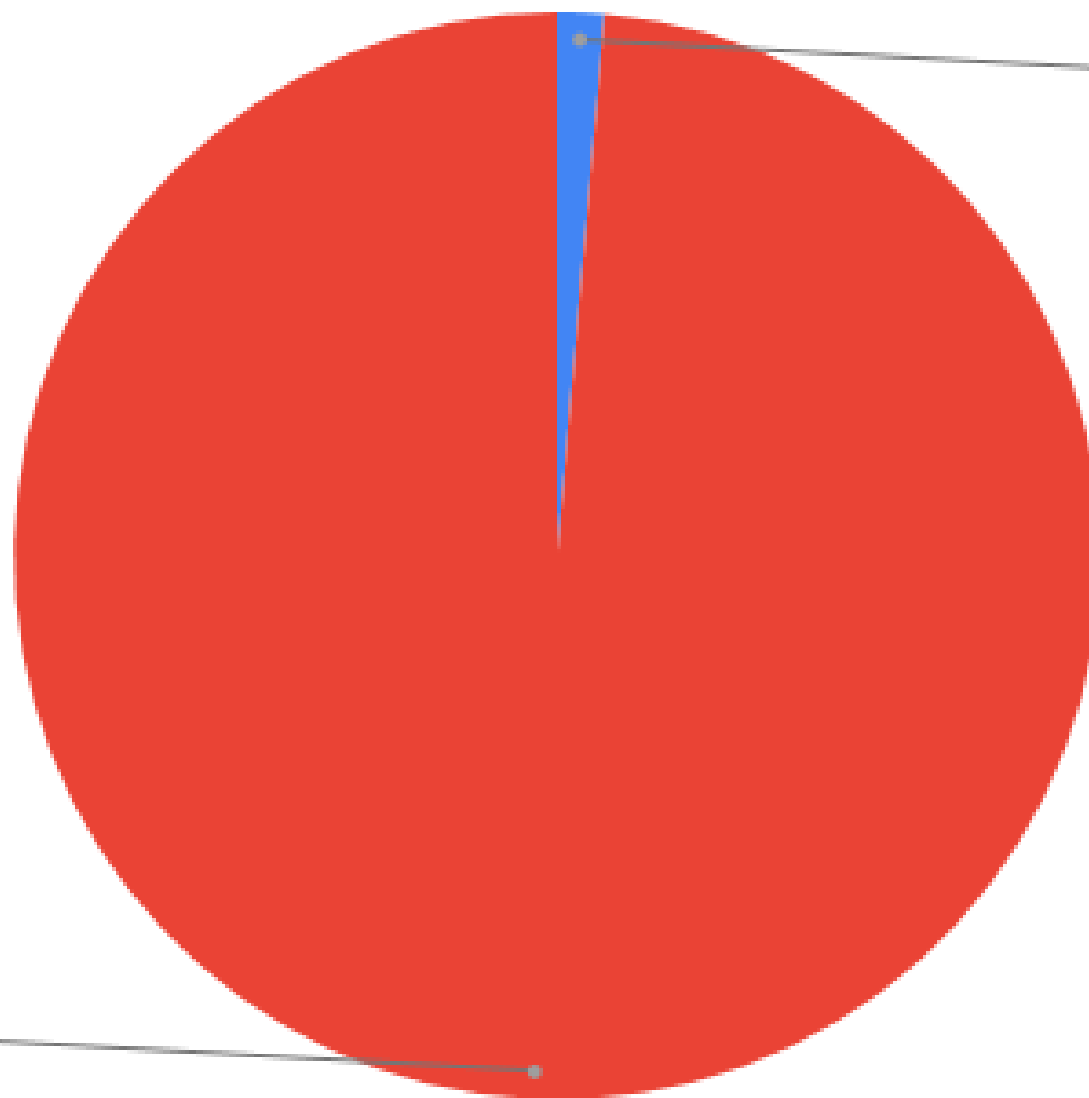
問題：
大腸桿菌？
所有物種？



含量

α

InterPro020476
98.6%



大腸桿菌
1.4%

3

Εύρηκα

避免取樣偏誤

所有已知

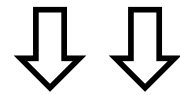


推論

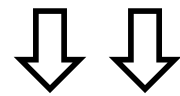
大腸桿菌



蒐集(手動挑選)
(InterPro)



資料處理(手動參數)
(EFI-EST)



可視化(手動染色)
(Cytoscape)

Εύρηκα



每一個點 = 一個蛋白質

每一條線 55% 相似

Low Statistical
Power

$\rho\eta\kappa\alpha$

β

mRNA-cap

二磷酸肌醇

輔酶A

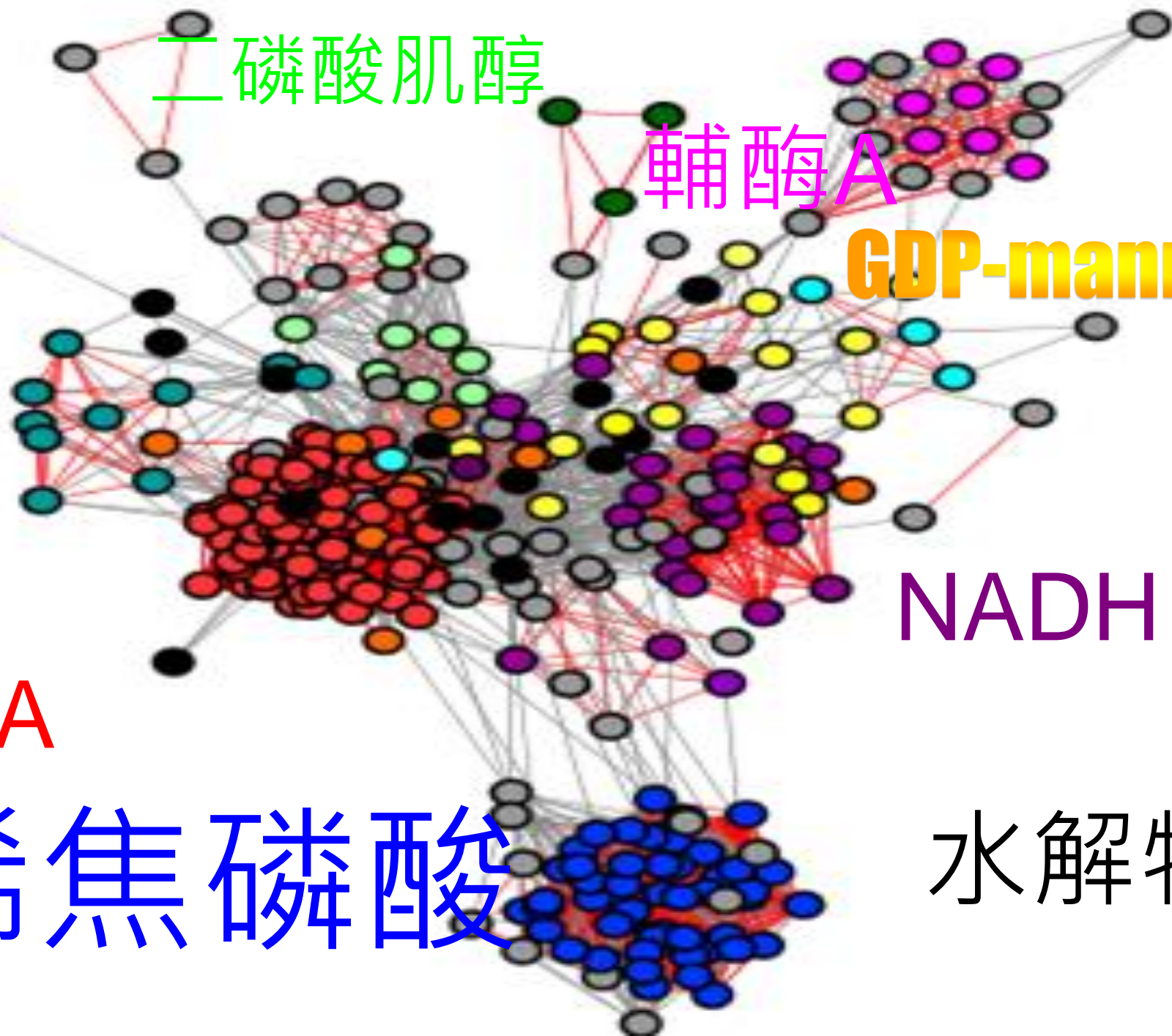
GDP-mannose

NADH

RNA

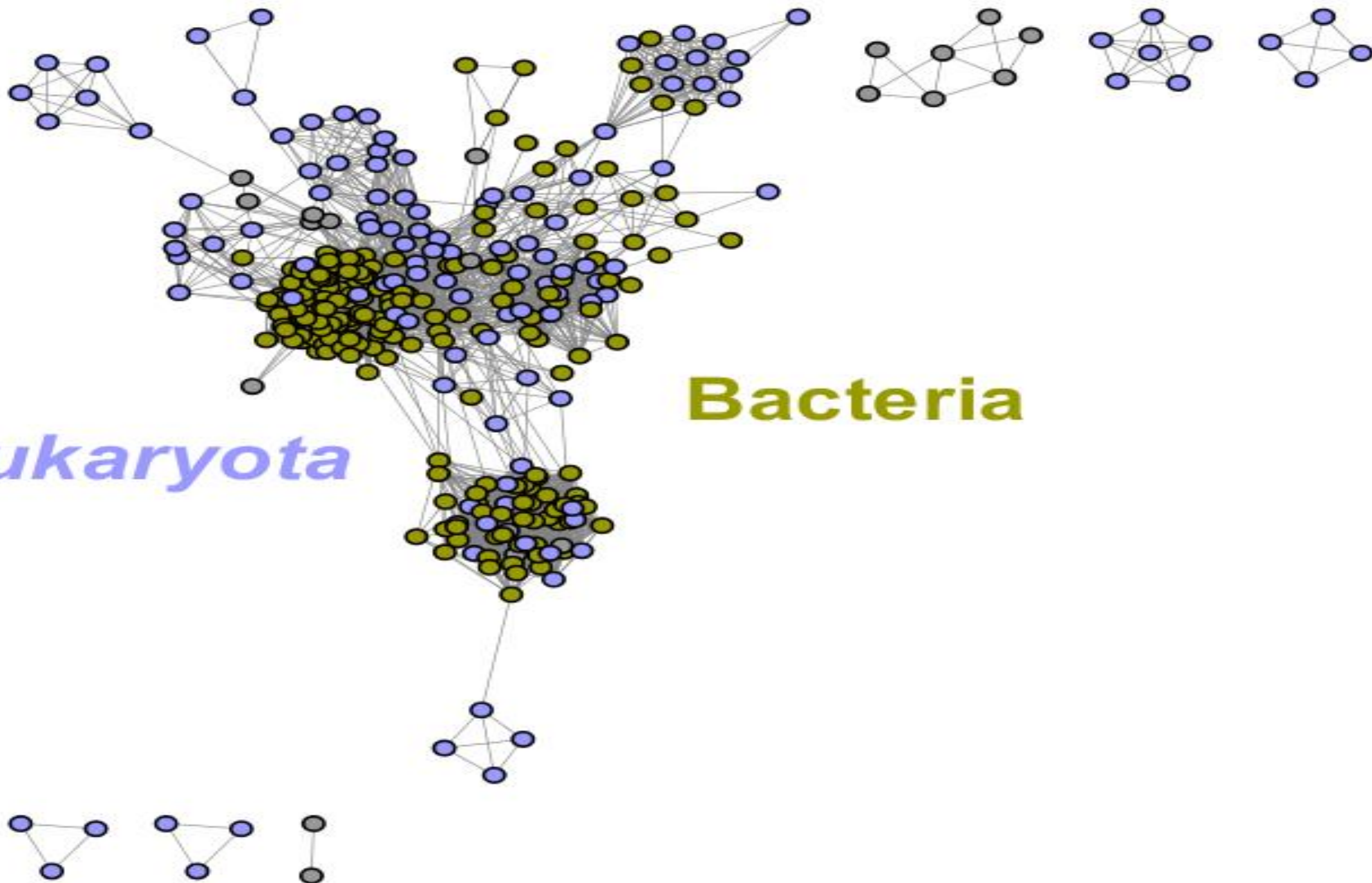
異戊烯焦磷酸

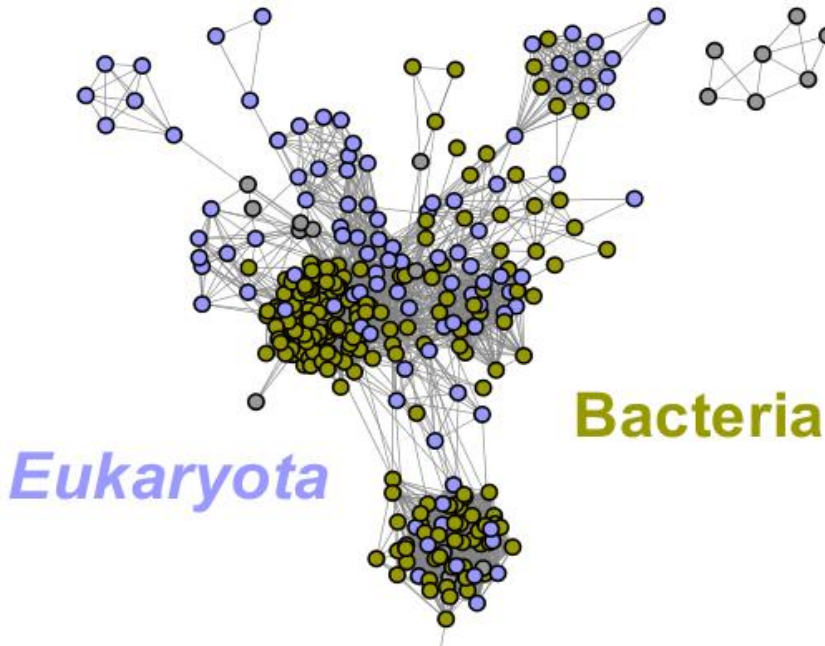
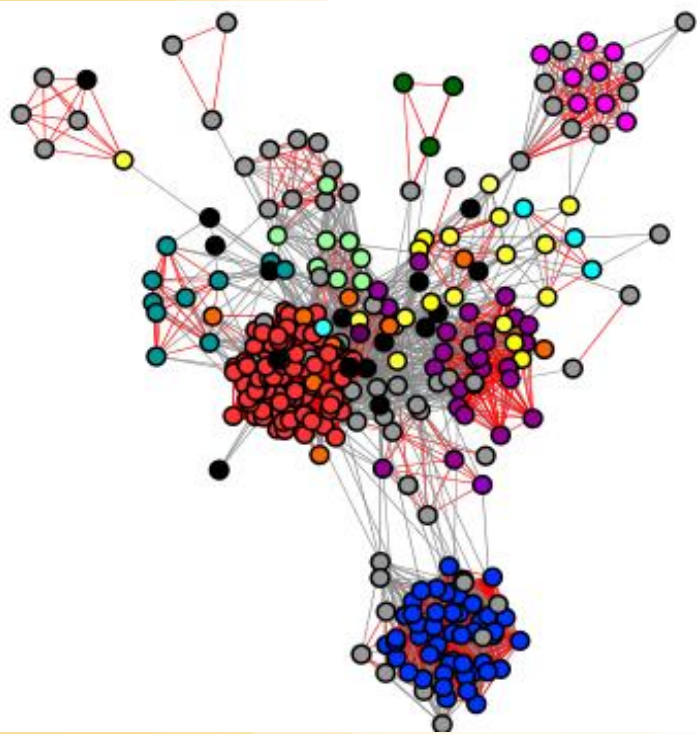
水解物質



Eukaryota

Bacteria





ρηκα

水解RNA為共祖蛋白最有可能的功能

一、演化趨向多樣化

1.紅色連接最多 (藍色<紅色)

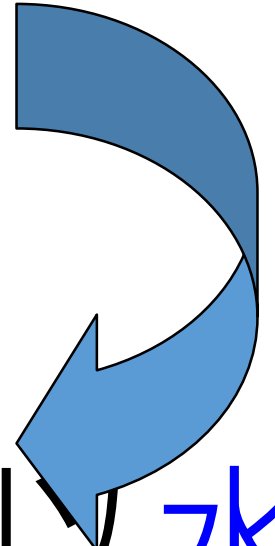
二、物種演化順序確定紅色非真核



Εύρηκα

水解RNA為共祖蛋
白最有可能的功能

為何有些NUDIX水解酶可以水
解超過一種核苷酸
分化過程中功能遺失



大腸桿菌

Εύρηκα

NudA (Mutt)	8-OH-dGTP
NudB	2-OH-dATP 、 dNTP
NudC	NADH 、 Ap ₂ A
NudD	GDP-mannose (sugar)
NudE	Ap ₃ A 、 ADP-ribose
NudF	ADP-sugar



大腸桿菌

Εύρηκα

NudG	5-OH-dCTP、2-OH-dATP、8-OH-dGTP
NudH	RNA
YeaB	未知
YmfB	TPP
YffH	GDP-mannose
YfaO	dUTP、dCTP、dTTP
YfcD	未知



Εύρηκα

定義四

Alignment
序列比對(排齊)

在最少刪減、延長下在不同序列
找到最多相似處的方法




```

      10           20           30           40           50
MTASAQPRGR RPGVGVGVVV TSKHPRCVL LGKRKGSVGA GSFQLPGGHL
      60           70           80           90          100
EFGETWEECA QRETWEEAAL HLKNVHFASV VNSFIEKENY HYVTILMKGE
      110          120          130          140          150
VDVTHDSEPK NVEPEKNESW EWVPWEELPP LDQLFWGLRC LKEQGYDPFK
      160
EDLNHLVG YK GNHL

```

Εύρηκα

```

      10           20           30           40           50
MYWSNQITRR LGERVQGFMG GISPQQMGEP EGWSGKNPG TMGASRLYTL
      60           70           80           90          100
VLVLQPQRVL LGMKKRGFGA GRWNGFGGKV QEGETIEDGA RRELQEESSL
      110          120          130          140          150
TVDALHKVGQ IVFEFVGEPE LMDVHVFCTD SIQGTPVESD EMRPCWFQLD
      160          170          180          190
QIPFKDMWPD DSYWFPLLQ KKKFHGYFEK QGQDTILDYT LREVDTV

```

```

sp|Q9NV35|NUD15_HUMAN -----MTASAQPRGRPGVGVG-----VVVTCKHPRCVL 30
sp|P36639|8ODP_HUMAN MYWSNQITRRLGERVQGFMSGISPQQMGEPGSGWSGKNPGTMGASRLYTLVLVLQPQRVL 60
                               .:*. * . * . . .:*. **

sp|Q9NV35|NUD15_HUMAN LGKRKGSVGAGSFQLPGGHLEFGETWEECAQRETWEEAALHLKNVHFAS-VVNSFIEKEN 89
sp|P36639|8ODP_HUMAN LGMKKRGFGAGRWNGFGGKVQEGETIEDGARRELQEESSLTVDALHKVGQIVFEFVGEPE 120
** :* ..*** :: **:: *** *: *:*** **:.* .: :* .. :* .*: : :

sp|Q9NV35|NUD15_HUMAN YHYVTILMKGEVDVTHDSEPKNVEPE-KNESW---EWVPWEELPPLDQLFWGLRCLK--E 143
sp|P36639|8ODP_HUMAN LMDVHVFCTDSIQ-----GTPVESDEMPCWFQLDQIPFKDMWPDDSYWFPLLQKKKF 174
      * :: ..... . ** : . .* : :*: : * * . : * *

sp|Q9NV35|NUD15_HUMAN QGYDPFKEDLNHLVG YKGNHL--- 164
sp|P36639|8ODP_HUMAN HG YFKFQGQ-DTILDYTLREVDTV 197
      **: * : : : :*: . .:

```

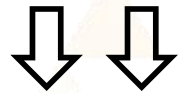


Εύρηκα

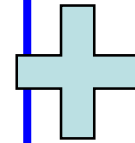
相異性高作比對會有
甚麼問題?



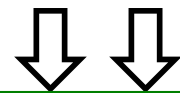
對於每一個物種，
同時具有不同功能的NUDIX



每個NUDIX是為
細胞功能而演化
而非跨物種的演化



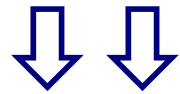
類似的序列意味
類似的功能



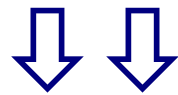
透過不同功能的蛋白質的序列(非物種)，
尋找其演化的趨勢

Εύρηκα

結構較序列演化上緩慢



結構更多相似處，比對更加準確



使用mTM-align



Εύρηκα



```

      cov  pid  1
1 mth1.pdb    100.0% 100.0% M-GASRLYTLVLVLQ---PQRVLLGMKKRGF-GAGRWNEFGGKVQEGETEDGARRELQEESELTVDALHKVGCQIVFEFV
2 mth2.pdb    86.5%  17.3% -GRRPGVGVGVVVTSCKHPRCVLLGKR-KGSVGAGSFQLPGGHLEFGTWECAQRETWEFAALHLKNVHFASVVNSFIE
consensus/100% ..ttstlhshVlVhp...PppVLLGh+.+G..GAGpaph.GG+lp.GEThe-sApREh.EEuuLplcslHhsu.ls..h.
consensus/90%  ..ttstlhshVlVhp...PppVLLGh+.+G..GAGpaph.GG+lp.GEThe-sApREh.EEuuLplcslHhsu.ls..h.
consensus/80%  ..ttstlhshVlVhp...PppVLLGh+.+G..GAGpaph.GG+lp.GEThe-sApREh.EEuuLplcslHhsu.ls..h.
consensus/70%  ..ttstlhshVlVhp...PppVLLGh+.+G..GAGpaph.GG+lp.GEThe-sApREh.EEuuLplcslHhsu.ls..h.

      cov  pid  81
1 mth1.pdb    100.0% 100.0% GEF---ELMDVHVFCTDSI--QS-TPVES-DE-M-RPCWFQLDQIP-FKDMWPDDSYWF--PILLQK-KK-FH--GYFKF
2 mth2.pdb    86.5%  17.3% ---KENYHYVTILMKGEVDVTHDSEPKNVEPEKNESWEEVWPWEELPPLDQLFWGLRCLKEQGYDPFKEDLNHLVGYKGN-
consensus/100% .....hhsshlhps-s...ps.pphps.sE...p.pWh.h-plP.hcpha.s.phhh..sh...K.ch.ah..hhhp.
consensus/90% .....hhsshlhps-s...ps.pphps.sE...p.pWh.h-plP.hcpha.s.phhh..sh...K.ch.ah..hhhp.
consensus/80% .....hhsshlhps-s...ps.pphps.sE...p.pWh.h-plP.hcpha.s.phhh..sh...K.ch.ah..hhhp.
consensus/70% .....hhsshlhps-s...ps.pphps.sE...p.pWh.h-plP.hcpha.s.phhh..sh...K.ch.ah..hhhp.

      cov  pid 161
1 mth1.pdb    100.0% 100.0% QCQDTILDYTREVDTV
2 mth2.pdb    86.5%  17.3% -----HL-----
consensus/100% .....ch.....
consensus/90% .....ch.....
consensus/80% .....ch.....
consensus/70% .....ch.....
```



	cov	pid
1 mth1.pdb	100.0%	100.0%
2 mth2.pdb	86.5%	17.3%
consensus/100%		
consensus/90%		
consensus/80%		
consensus/70%		

```

1 [
M-GASRLYTLVLVLQ---PQRVLLGMKKRSE-GAGRWNGFGGKVQEGETIEDGARRELQEESGLTVDALHKVGQIVFEFV
-GRRPGVGVGVVVVTSCKHPRCVLLGKR-KESVGAGSFQLPGGHLEFGETWEECAQRETWEEAALHLKNVHFASVVSNSIE
..ttstlhshVlVhp...PppVLLGh+.+G..GAGpaph.GG+lp.GETHe-sApREh.EEuuLplcslHhsu.ls..h.
..ttstlhshVlVhp...PppVLLGh+.+G..GAGpaph.GG+lp.GETHe-sApREh.EEuuLplcslHhsu.ls..h.
..ttstlhshVlVhp...PppVLLGh+.+G..GAGpaph.GG+lp.GETHe-sApREh.EEuuLplcslHhsu.ls..h.
..ttstlhshVlVhp...PppVLLGh+.+G..GAGpaph.GG+lp.GETHe-sApREh.EEuuLplcslHhsu.ls..h.

```

pnka

	cov	pid
1 mth1.pdb	100.0%	100.0%
2 mth2.pdb	86.5%	17.3%
consensus/100%		
consensus/90%		
consensus/80%		
consensus/70%		

```

81
GEP---ELMDVHVFCTDSI--QG-TPVES-DE-M-RPCWFQDQIP-FKDMWPDDSYWF--PLLLQK-KK-FH--GYFKF
---KENYHYVTILMKGEVDVTHDSEPKNVEPEKNESWEWVPWEELPPLDQLFWGLRCLKEQGYDPFKEDLNHLVGYKGN-
.....hhsshlhps-s...ps.pPhps.sE...p.pWh.h-plP.hcpha.s.phhh..sh...K.ch.ah..hhhp.
.....hhsshlhps-s...ps.pPhps.sE...p.pWh.h-plP.hcpha.s.phhh..sh...K.ch.ah..hhhp.
.....hhsshlhps-s...ps.pPhps.sE...p.pWh.h-plP.hcpha.s.phhh..sh...K.ch.ah..hhhp.
.....hhsshlhps-s...ps.pPhps.sE...p.pWh.h-plP.hcpha.s.phhh..sh...K.ch.ah..hhhp.

```

	cov	pid
1 mth1.pdb	100.0%	100.0%
2 mth2.pdb	86.5%	17.3%
consensus/100%		
consensus/90%		
consensus/80%		
consensus/70%		

```

161
QGQDTILDYTLREVDTV
-----HL-----
.....ch.....
.....ch.....
.....ch.....
.....ch.....

```

GAP(-) 比較少

sp|Q9NV35|NUD15_HUMAN
sp|P36639|8ODP_HUMAN

```

-----MTASAQPRGRPGVGVG-----VVVTSCKHPRCVL 30
MYWSNQITRRLGERVQGFMSGISPPQMGEPGSGWSGKNPGTMGASRLYTLVLVLQPQRVL 60
.:*. * . * . . . :*: **

```

sp|Q9NV35|NUD15_HUMAN
sp|P36639|8ODP_HUMAN

```

LGKRKGSVAGSGFQLPGGHLEFGETWEECAQRETWEEAALHLKNVHFAS-VVNSFIEKEN 89
LGMKKRGFGAGRWNGFGGKVQEGETIEDGARRELQEESGLTVDALHKVGQIVFEFVGEPE 120
** :* ..*** :: **::: *** *: *:*** **:.* .. :* .. :* :*: :

```

sp|Q9NV35|NUD15_HUMAN
sp|P36639|8ODP_HUMAN

```

YHYVTILMKGEVDVTHDSEPKNVEPE-KNESW---EWVPWEELPPLDQLFWGLRCLK--E 143
LMDVHVFCTDSIQ-----GTPVESDEMPCWFQDQIPFKDMWPDDSYWFPLLLQKKKF 174
* :: :.: : . ** : . .* : :*: : * *. : : *

```

sp|Q9NV35|NUD15_HUMAN
sp|P36639|8ODP_HUMAN

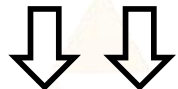
```

QGYDPFKEDLNHLVGYKGNHL--- 164
HGYFKFQGQ-DTILDYTLREVDTV 197
:** *: : : :.:* . .

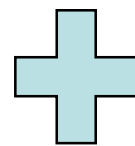
```

B

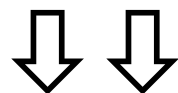
對於每一個物種，
同時具有不同功能的NUDIX



每個NUDIX是為
細胞功能而演化
而非跨物種的演化

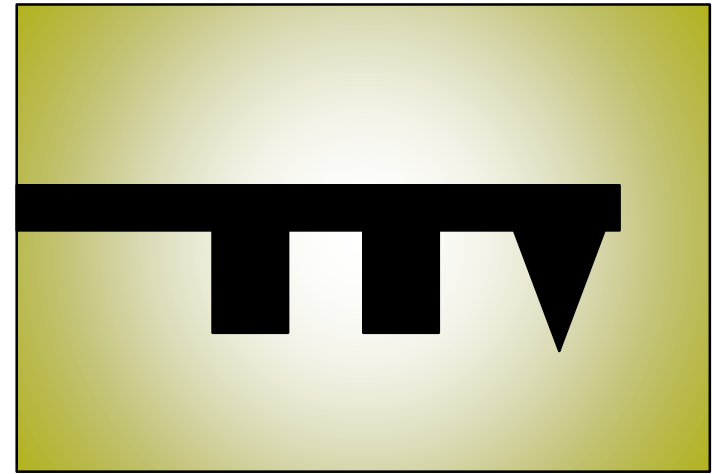
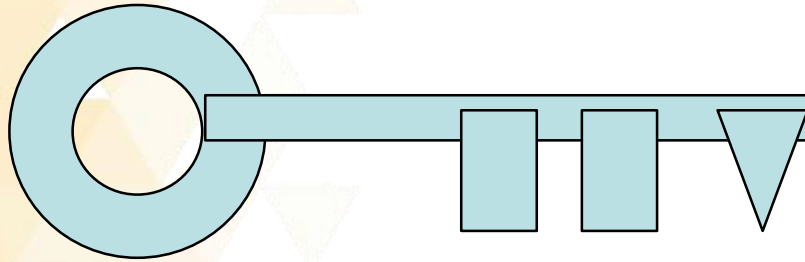


類似的結構
更加意味類
似的功能



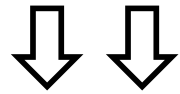
透過不同功能的蛋白質的序列(非物種)，
尋找其演化的趨勢

Εύρηκα



類似的鑰匙孔更加意味類似的鑰匙

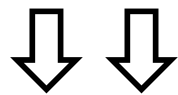
蒐集(手動挑選12個結構序列)
(PDBeBank)



資料處理

做序列分析 (mTM-align)

做演化樹分析 (PhyML)



可視化
(IToL)

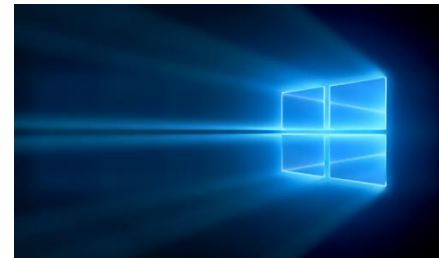
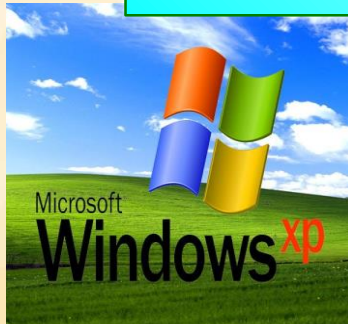


Maximum-Likelihood的演化假設

哪一個演化樹容易發生？



永遠選後設機率較高者



cons

cons

cons

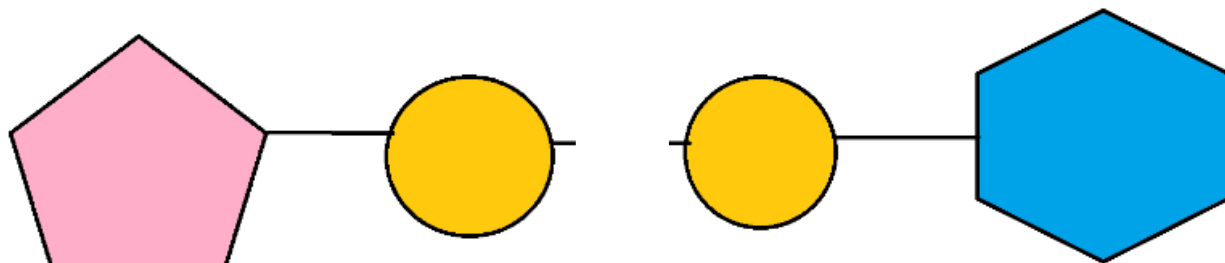
CONS

cons

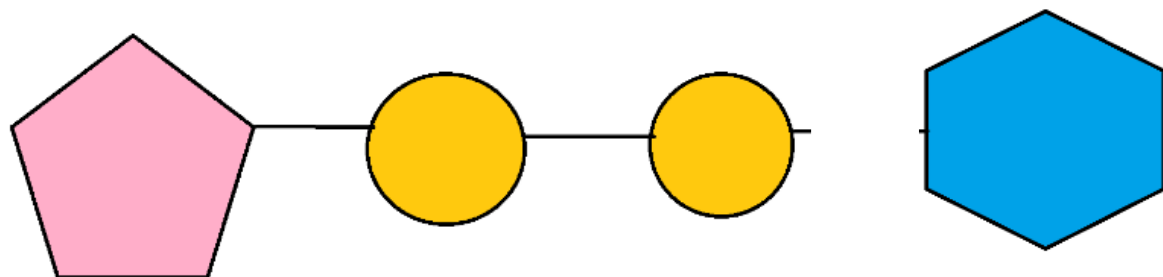
cons

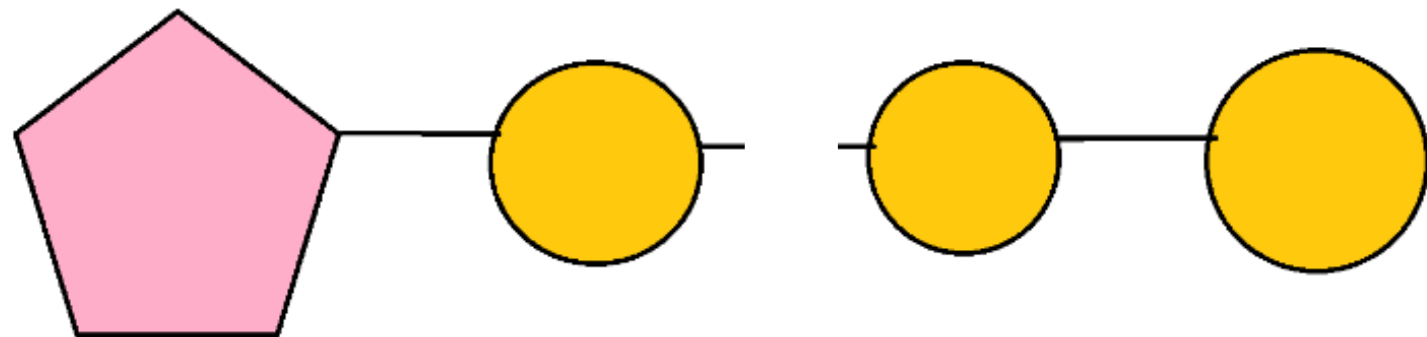
Εύρηκα

第一種

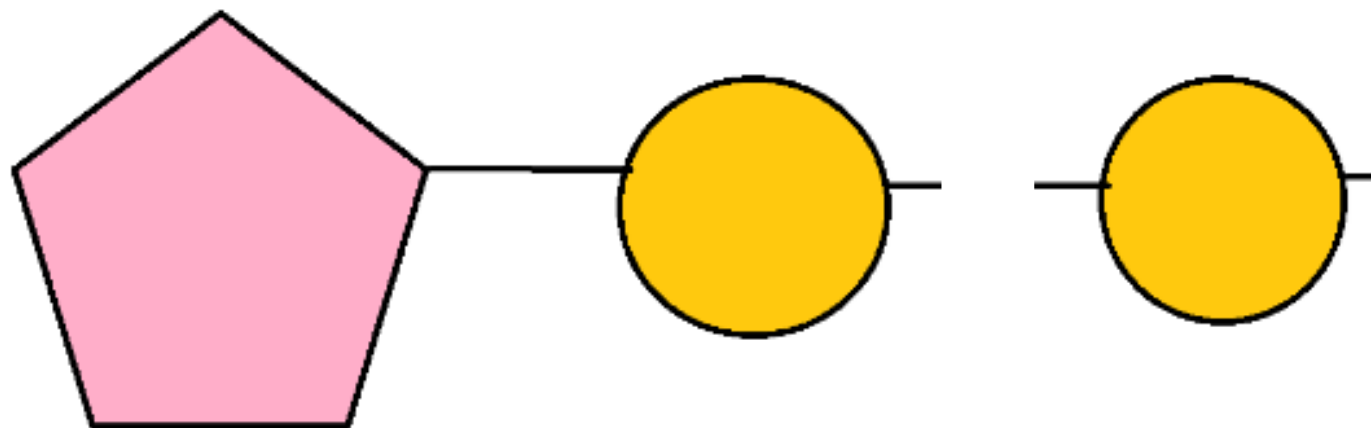


第二種





δρηνκα
第三種



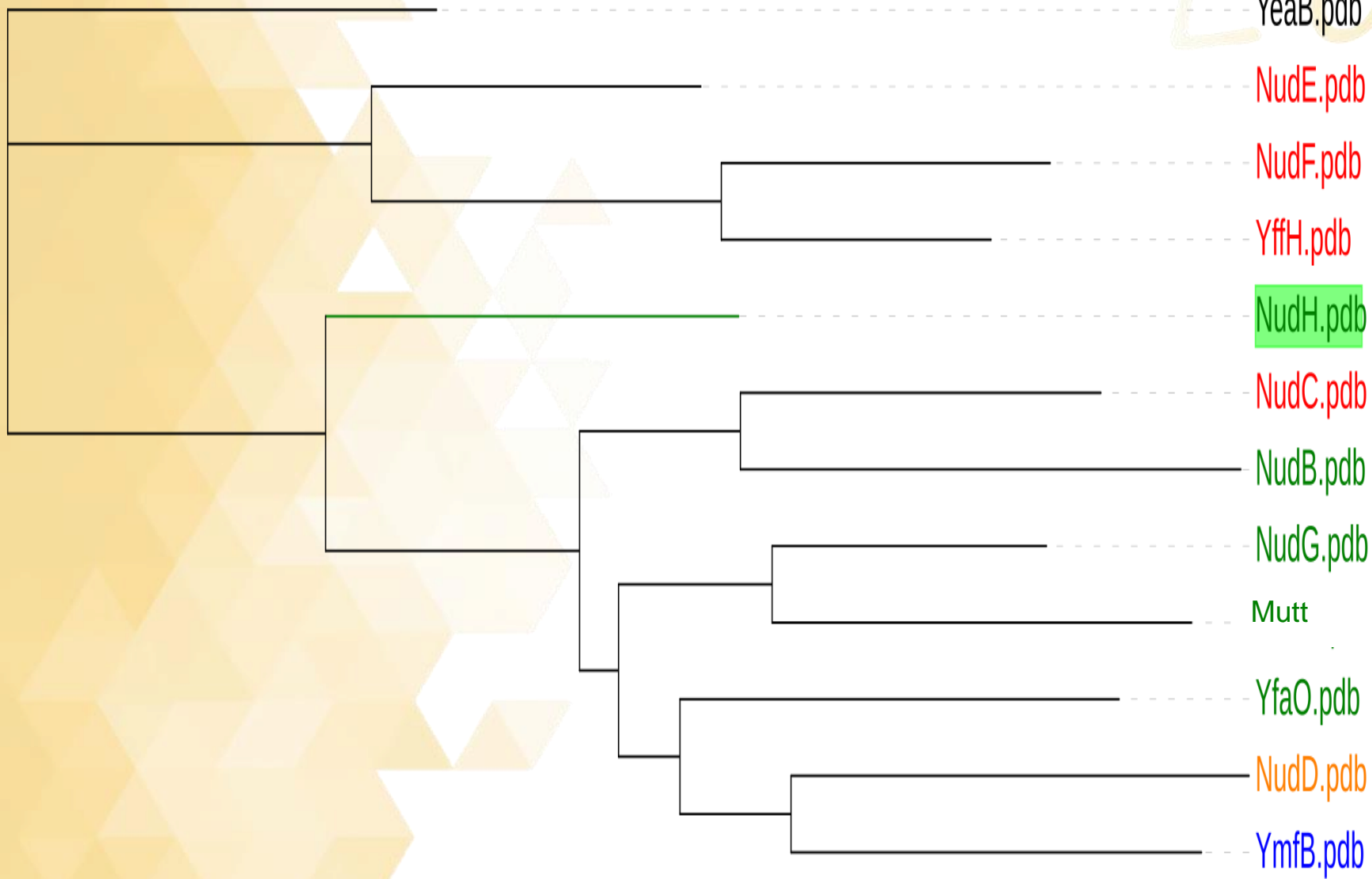
第四種

Εύρηκα

Structure Only



Ευρηκα

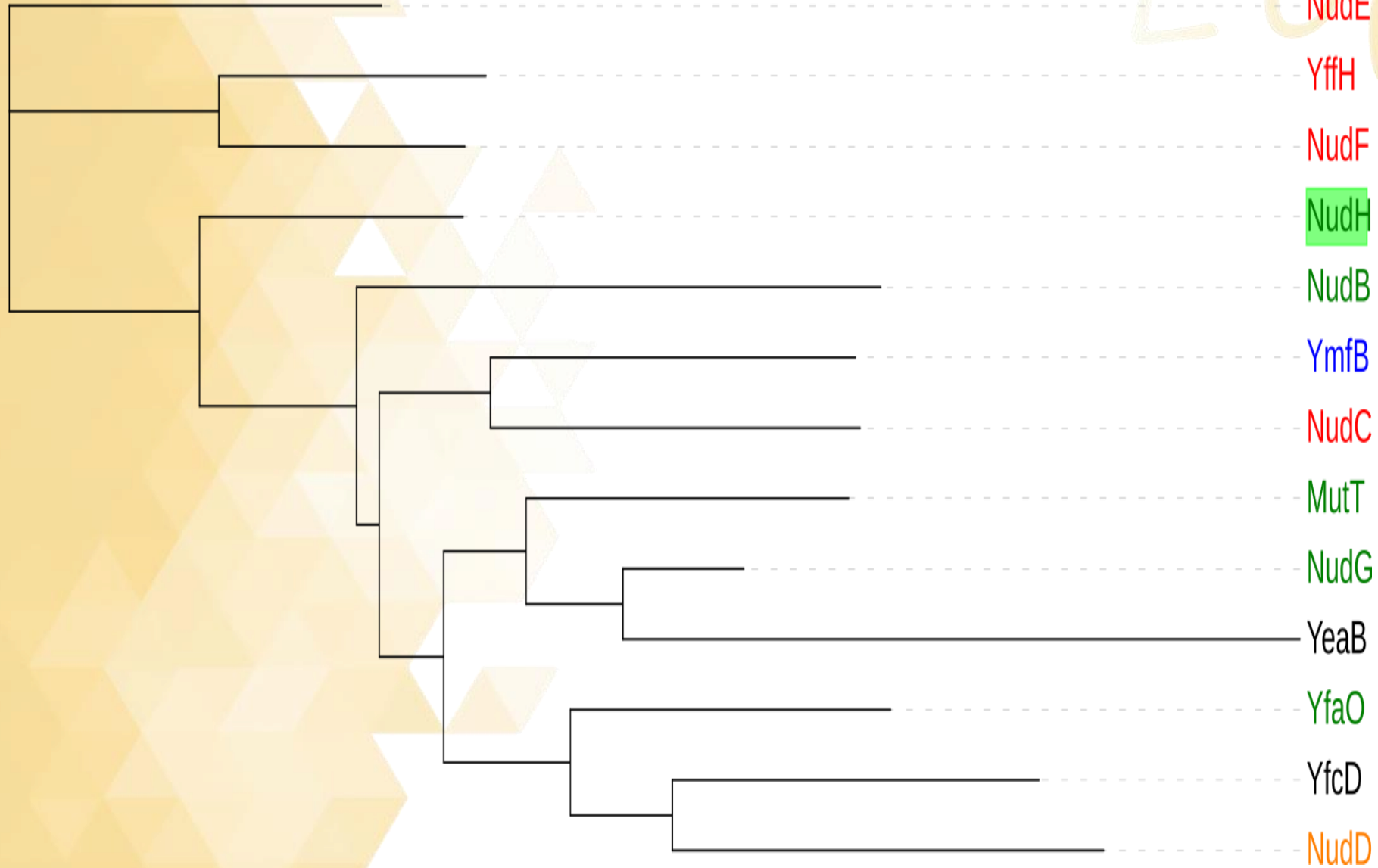


Εύρηκα

Sequence Only



Εύρηκα



Εύρηκα

NudA
NudG



Εύρηκα

NudA (Mutt)	8-OH-dGTP
NudB	2-OH-dATP 、 dNTP
NudC	NADH 、 Ap ₂ A
NudD	GDP-mannose (sugar)
NudE	Ap ₃ A 、 ADP-ribose
NudF	ADP-sugar



NudG	5-OH-dCTP 、 2-OH-dATP 、 8-OH-dGTP
NudH	RNA
YeaB	未知
YmfB	TPP
YffH	GDP-mannose
YfaO	dUTP 、 dCTP 、 dTTP
YfcD	未知



Εύρηκα

NudE

NudF

YffH



NudA (Mutt)	8-OH-dGTP
NudB	2-OH-dATP 、 dNTP
NudC	NADH 、 Ap ₂ A
NudD	GDP-mannose (sugar)
NudE	Ap ₃ A 、 ADP-ribose
NudF	ADP-sugar



NudG	5-OH-dCTP 、 2-OH-dATP 、 8-OH-dGTP
NudH	RNA
YeaB	未知
YmfB	TPP
YffH	GDP-mannose
YfaO	dUTP 、 dCTP 、 dTTP
YfcD	未知



NudA (Mutt)	8-OH-dGTP
NudB	2-OH-dATP 、 dNTP
NudC	NADH 、 Ap ₂ A
NudD	GDP-mannose (sugar)
NudE	Ap ₃ A 、 ADP-ribose
NudF	ADP-sugar



Εύρηκα

是否存在錯誤？

YffH 、 NudD看似功能相似
卻不被演化樹分在一起。

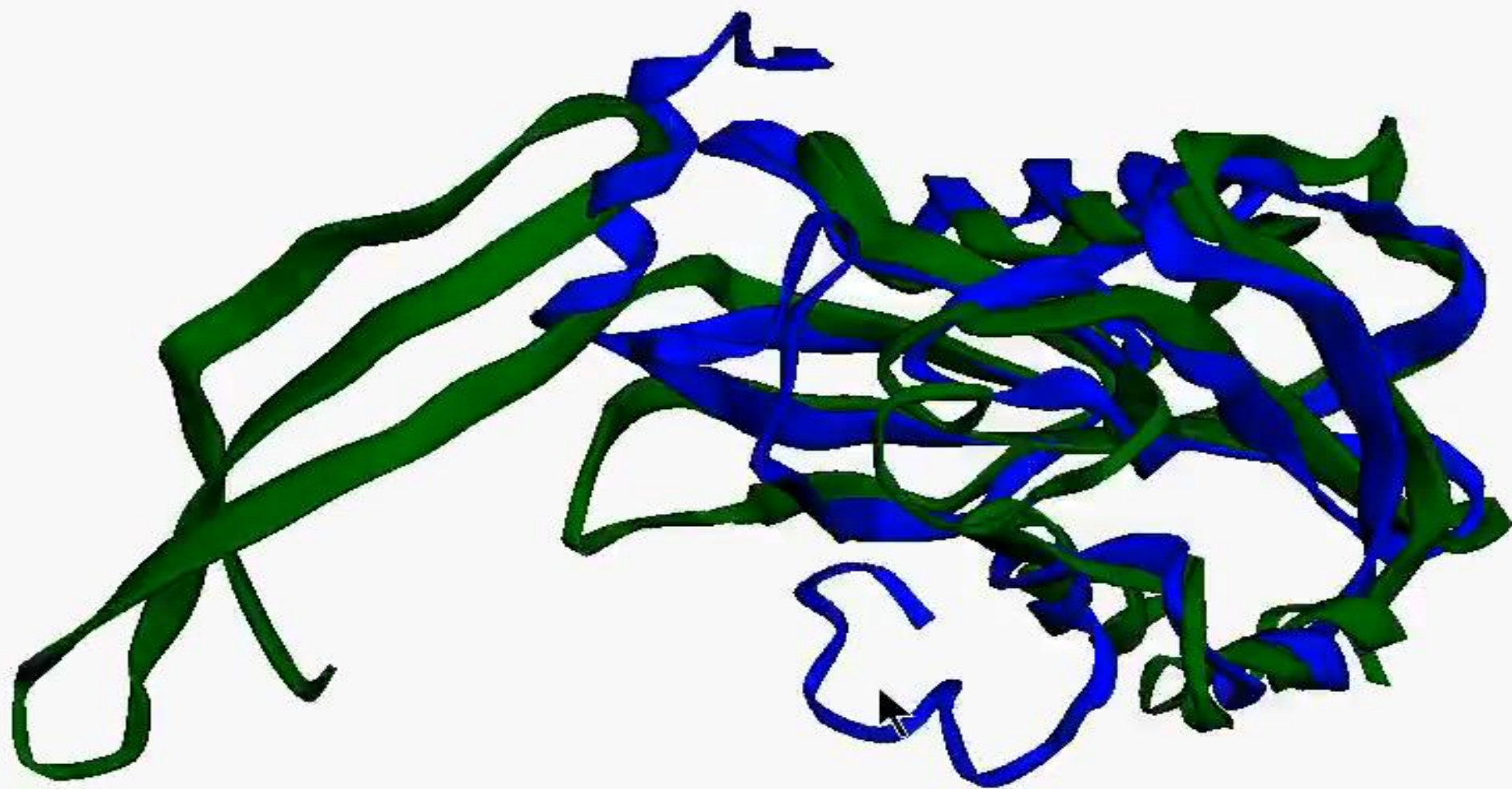


Εύρηκα

NudD

YffH



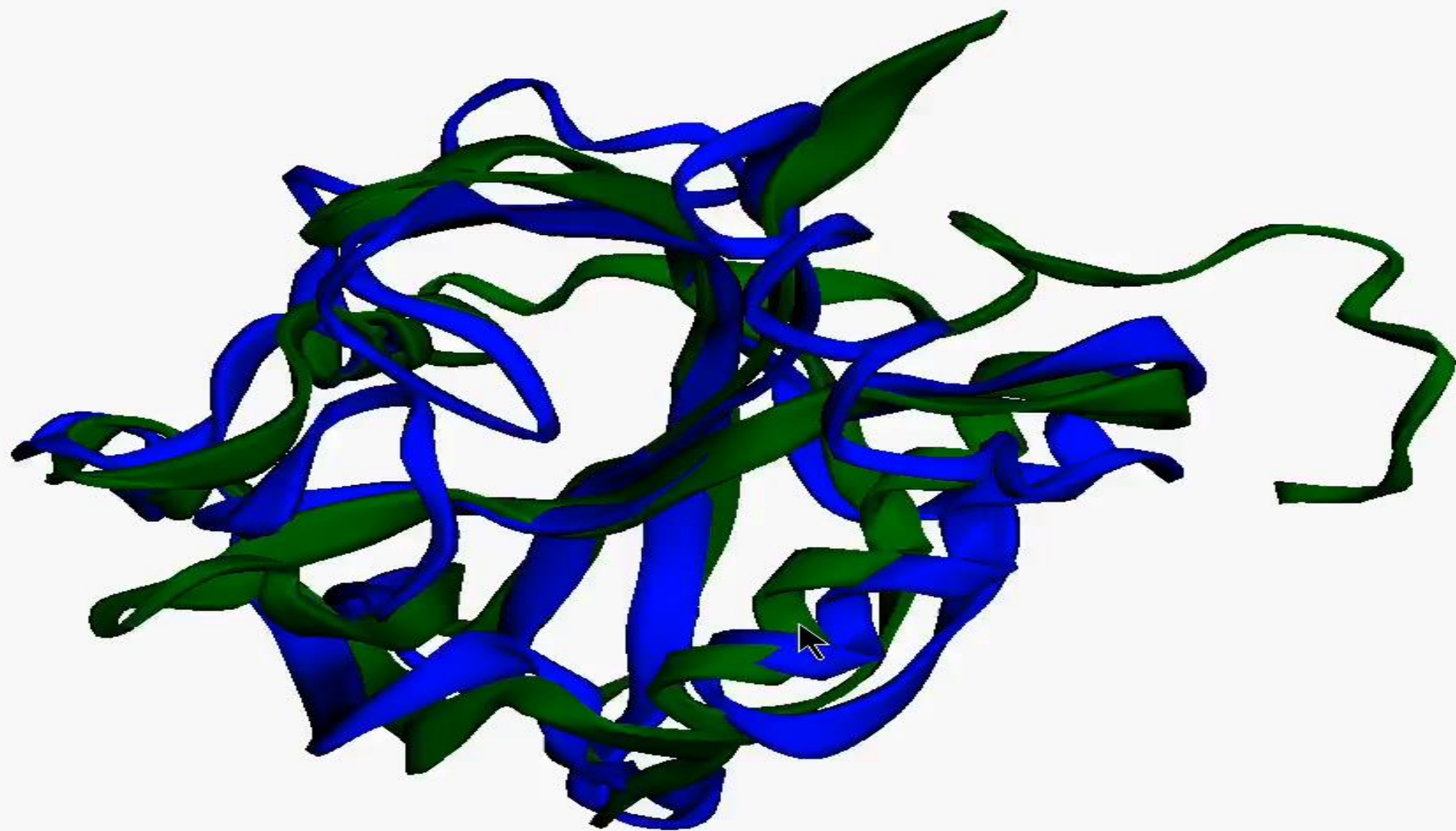


Εύρηκα

NudD

YmfB





假設

Εύρηκα

- 一、演化趨向多樣化
- 二、每個NUDIX是為細胞功能而演化
- 三、結構較序列演化上緩慢
- 四、類似的結構更加意味類似的功能
- 五、演化樹永遠選後設機率較高者



結論

Εύρηκα

- 一、水解RNA為共祖蛋白最有可能的功能
- 二、比對發現功能十分相似的NUDIX水解酶



未來展望

Εύρηκα

- 一、催化的點在序列的何處? (功能性)
- 二、等到有更多資料再畫一次SSN



感謝鄭貽生 教授
提供研究方向與器材

Εύρηκα

實驗室：台大生科館1121室

特色： MTH1----癌症治療曙光

纖維素合成酶----生質能源之希冀

植物去乙醯基酶----植物對基因的調控

優勢：可以利用結構來了解功能！



Εύρηκα

謝謝各位的時間
與鼓勵!

感激不盡!

