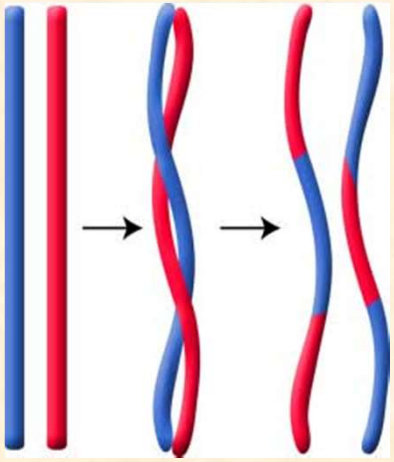
*http://petang.cgu.edu.tw/bioinformatics/index.htm*



# Bioinformatics

## *Lecture 3 : Bio-computing and Linux Basics*

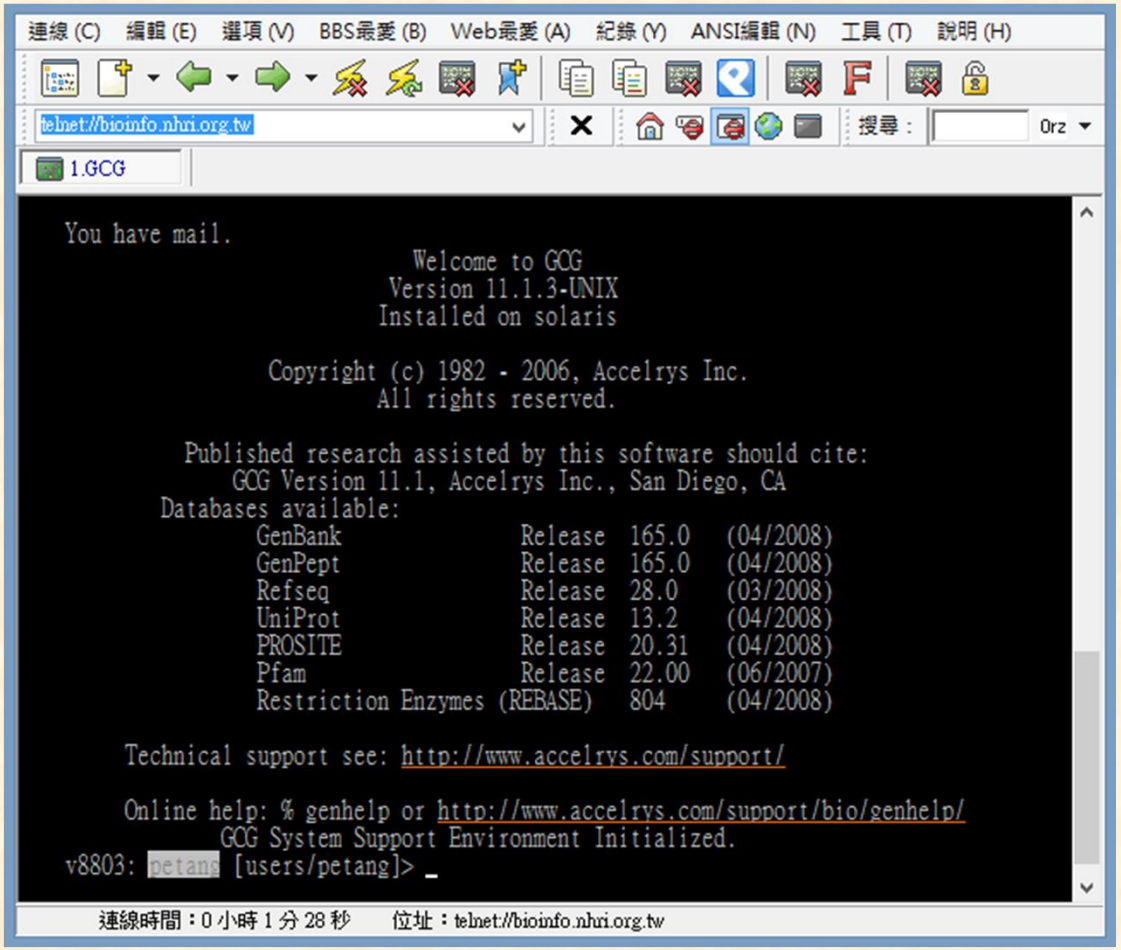
AGTCCGCGAATACAGGCTCGGT AGTCCGCGAATACAGGCTCGGT

**Petrus Tang, Ph.D. (**鄧致剛**)**
**Graduate Institute of Basic Medical Sciences**
**and**
**Bioinformatics Center, Chang Gung University.**
**petang@mail.cgu.edu.tw**
**EXT:** 5136

助教：
蔡智宇(分機5690)

# UNIX SYSTEM

UNIX 是個多人多工作業系統。另外，UNIX 多種不同之版本如 SunOS、Linux 等等。但各系統之基本操作指令大同小異。



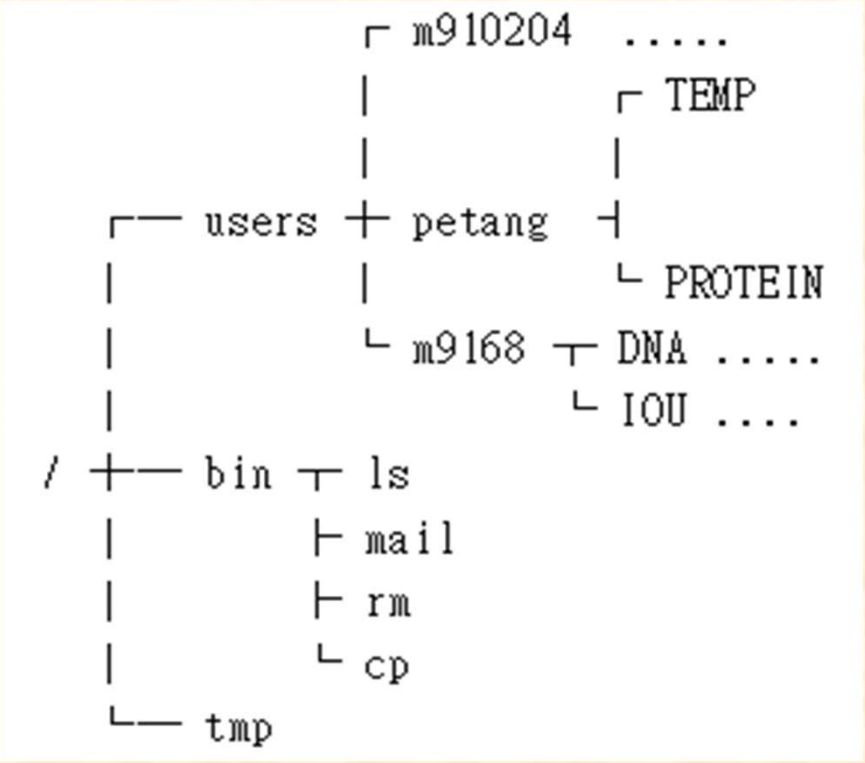
# GCG UNIX 作業系統操作簡介-1

## 檔案、目錄、與檔案系統

■ 每一個使用者在 UNIX 下面均有一個 HOME directory ，做為這個使用者個人資料的存放處。使用者 login 時，系統管理者設定進入工作目錄的位置。

■ UNIX 使用一個和 DOS 相似的 single root 樹狀結構檔案系統，由根目錄( / )開始一層層將子目錄建下去，各階層目錄以 / 隔開。

■ 最頂層稱為根目錄( root )，包含許多目錄，每一個目錄底下均存放許多的檔案與子目錄。



# GCG UNIX 作業系統操作簡介-2

## 指令格式

指令 [參數] [處理對象]
gcg2% [command] [options] [file]
例：ls –l aminoacid.dat
　　除了列出檔名外，並列出檔案屬性及擁有者、檔案大小及建立時間等資訊

- 在 UNIX 之下英文字母的大小寫是有差別的，命令一般是小寫字串。
  本課程建議所有"根目錄"以大寫命名。其它檔案全部用小寫。
- 選項通常以減號(-)再加上一個或數個字元表示，
  用來選擇一個命令的不同操作
- 同一行可下數個命令，命令間應以分號隔開。
- 線上求助指令 -- man　可在線上用來查詢各種命令用法(manual page)的指令
  例：man ls　查詢 ls 這個指令的用法

# GCG UNIX 作業系統操作簡介-3




基本指令

- **ls** list files
- **more** display contents of file on screen page by page
- **cat** display contents the whole file
- **cp** make a new copy of a file
- **rm** remove (delete) a file
- **mv** rename (move) a file
- **mkdir** make a new directory
- **rmdir** remove a directory

# GCG UNIX 基本指令 - ls

ls 的意義為 "list"，也就是列出目錄內檔案名稱（如 DOS 的 dir/w）

- 如果你在下 ls 指令後頭沒有跟著任何的檔名，
  它將會顯示出目前目錄中所有檔案。
- 也可以在 ls 後面加上所要察看的目錄名稱或檔案的名稱，如

```
gcg2% ls filename
gcg3% ls ROOT
```

ls -[參數]

ls -l 除了列出檔名外，並列出檔案屬性及擁有者、檔案大小及建立時間等資訊
ls -a 列出所有檔案，包括隱藏檔
ls -R 地列出所有子目錄內所有檔案

ls -[參數] [filename 萬用字元(wildcard)]

萬用字元有三種：

* 代表任意長度的字串
? 代表任何的單一字元
[...] 代表區間內定義的任一字元

c* 代表cat ，cp ，cal ...任意長度且開頭為c 。
ca? 代表cat，cal ...三個字且前二字為 ca。
ca[a-z] 代表caa，cab，caz ... 三個字且頭二字為 ca 第三字為 a-z。

# GCG UNIX 基本指令 – cat/file

以上二個指令均為察看檔案內容的指令。

- cat [filename]

是 concatenate ，在字典上的意思是“連結,將…串成鎖狀” 。就是把檔案的內容顯示出來的意思。
問題在於 cat 它是不會停下來的，因此並不好用( 試想如果一個螢幕二十四行，而一個檔案四百行，cat 一出來將會劈哩啪啦不斷的捲上去，使用者很難據此得到他們所需的資訊。
cat 可配合萬用字元使用。

- more [filename]

可以將所觀察的檔案跟據終端機的形態一頁頁的顯示出來，
再根據使用者的要求換頁或捲行。
如果你在使用中覺得已經看到了所要看的部份，
可以按 'q‘ 離開 more 的使用。

# GCG UNIX 基本指令 - cp

這個指令的意義是複製 ("COPY") ,
也就是將一個或多個檔案複製成另一個檔案或者是將其複製到另一個目錄。

- cp 的用法如下:

cp f1 f2 : 將檔名為 f1 的檔案複製一份為檔名為 f2 的檔案。
cp f1 f2 f3 ... dir : 將檔案 f1 f2 f3 ... 都以相同的檔名複製一份放到目錄 dir 裡面。
cp -r dir1 dir2 : 將 dir1 的全部內容全部複製到 dir2 裡面。

- cp -[參數]:

-i : 此參數是當已有檔名為 f2 的檔案時，若逕自使用 cp 將會將原來 f2 的內容掩蓋過去，因此在要蓋過之前必須先詢問使用者一下。如使用者的回答是y(yes)才執行複製的動作。

-r : 此參數是用來做遞迴複製用，可將一整顆子樹都複製到另一個目錄中。

# GCG UNIX 基本指令 - mv

mv 的意義為 move , 主要是將一檔案改名或換至另一個目錄。
如同 cp ，它也有三種格式:

mv f1 f2 : 將檔名為 f1 的檔案變更成檔名為 f2 的檔案。
mv dir1 dir2 : 將檔名為 dir1 的目錄變更成檔名為 dir2 的目錄。
mv f1 f2 f3 ... dir : 將檔案 f1 f2 f3 ... 都移至目錄 dir 裡面。

■ mv-[參數]
-i：此參數的意義與 cp 中的相同，均是 interactive 詢問之意。
-f：此參數為強迫( force ) , 就是不管有沒有同名的檔案，反正我就是要搬過去，所有其他的參數遇到 -f 均會失效。

# GCG UNIX 基本指令 - rm

rm 的意義是 remove ，也就是用來殺掉一個檔案的指令。

在 UNIX 中一個被殺掉的檔案除非是系統恰好有做備份，否則是無法像 DOS 裡面一樣還能夠救回來的。所以在做 rm 動作的時要特別小心。

rm 的格式如下:
rm  f1 f2 f3 .....

- rm-[參數]
  - -f : 將會使得系統在刪除時，不提出任何警告訊息。
  - -i : 在除去檔案之前均會詢問是否真要除去。
  - -r : 遞迴式的刪除。

# GCG UNIX 基本指令 – mkdir/rmdir

- mkdir 是一個讓使用者建立一個目錄的指令。

你可以在一個目錄底下使用 midir 建立一個子目錄，使用的方法如下:

mkdir dirname1 [ dirname2 ... ]

如此你就可以建立一個或多個目錄

- 相對於 mkdir ，rmdir 是用來將一個"空的"目錄殺掉的。如果一個目錄下面沒有任何檔案，你就可以用 rmdir 指令將其除去。rmdir 的使用法如下:

rmdir dirname1 [ dirname2 .... ]

如果一個目錄底下有其他的檔案， rmdir 將無法將這個目錄殺掉，除非使用 rm 指令的 -r 選項。

# GCG UNIX 終端機常用控制鍵

Ctrl-C　中斷程式的執行。

Ctrl-Z　暫停程式的執行，稍後可下 fg 或 bg 指令繼續，若未下 fg 或 bg 指令繼續執行，該 process 仍會留在系統內。

# EXERCISE 


**Exercise 02-1 Linux Commands (download file Ex from website)**

- *learn how to use basic linux commands for file management*

Exercise 02-2 File Transfer via FTP

- *learn how import and export files to/from GCG*

Exercise 02-3 Sequence Manipulation in GCG UNIX

Exercise 02-4 Sequence Manipulation in GCG UNIX : LEGUMAIN

http://pcman.ptt.cc/




http://pcman.ptt.cc/windows.html

*http://pcman.ptt.cc/windows.html*



# PCMan BBS Project

開放原始碼的瀏覽 BBS 解決方案

首頁　下載　開發　關於

## PCMan Novus

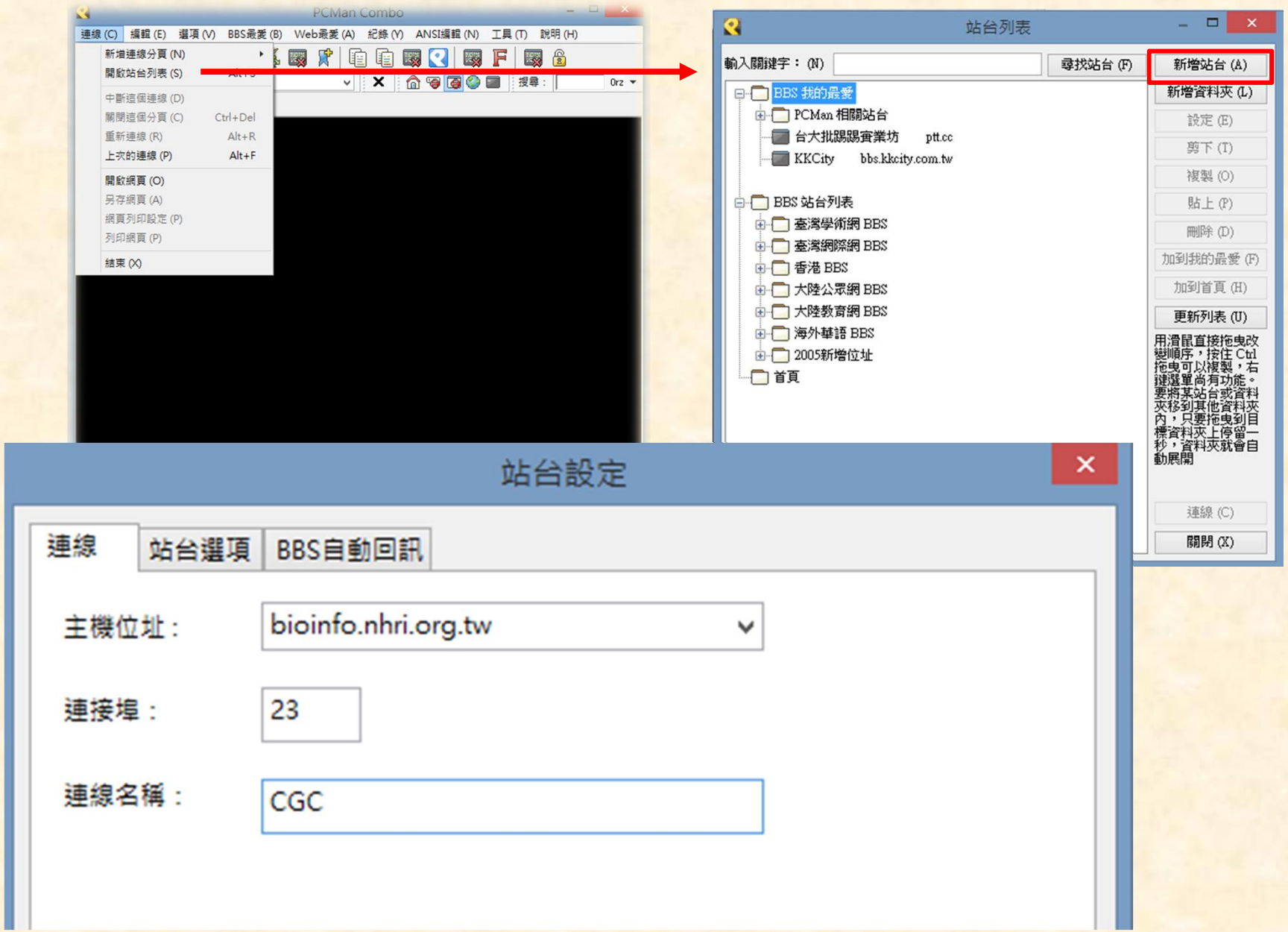
Windows 平台上的系列元祖

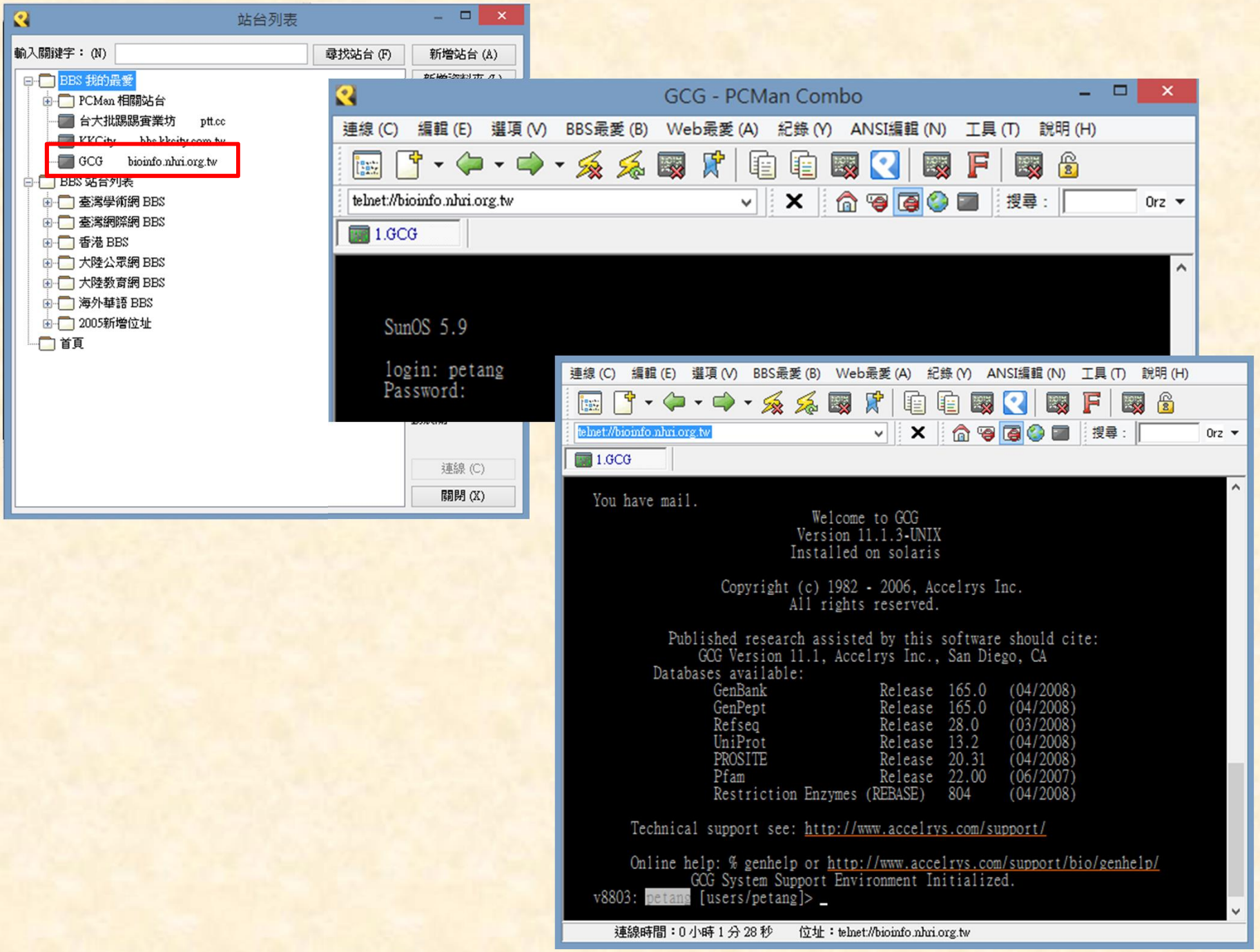
### 最新版本

- PCMan Lite Novus Ver. 9.4.2
- PCMan Combo Novus Ver. 9.4.2

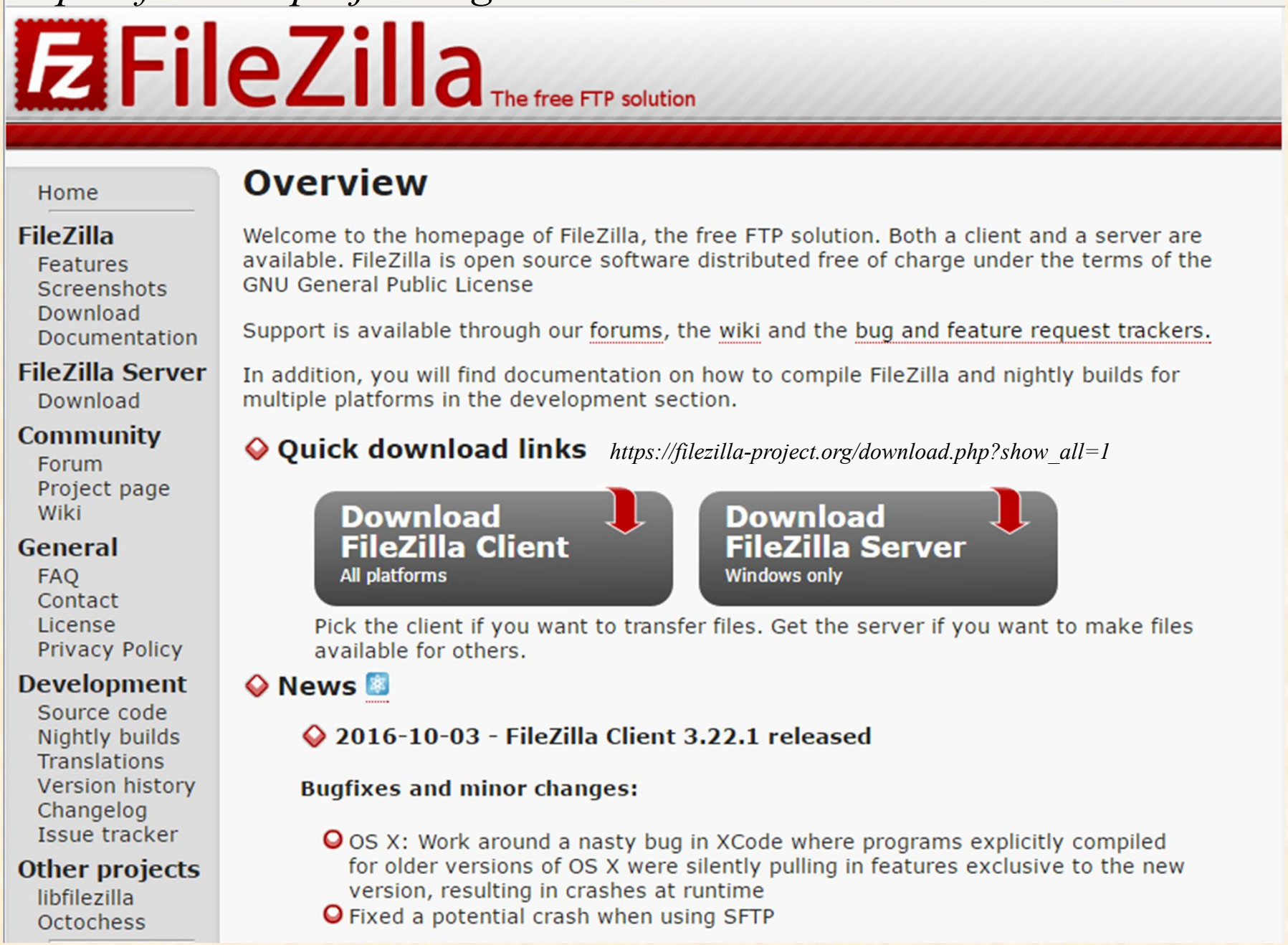
### 主要特色

- 國產免費自由軟體，中文支援度 100%，不含廣告，且原始碼全部以 GNU GPL 授權開放
- 體積輕巧、啟動快速，節省系統資源！ (連 Pentium II 266 MHz 的老電腦都跑得動)
- 同時提供支援瀏覽器的 Web/BBS 兩用版本，以及純 BBS 版本，符合多樣化需求
- 使用分頁瀏覽，開啟多重視窗，同時瀏覽多站台最輕鬆
- 內建 Unicode 補完，可顯示及輸入 Big 5 日文，不需安裝 UAO (Unicode 補完計畫)
- 內建繁簡體中文輸入/顯示轉換功能 (Big5 <-> GB2312)，瀏覽兩岸三地 BBS 站沒困擾
- 所有操作熱鍵、工具列按鈕、圖示皆可自訂，符合使用者習慣，操作最方便
- 支援 ANSI 彩色，並內建彩色文字編輯器，可自由編輯、複製、貼上 BBS 彩色文字
- 內建多種可愛表情符號，可快速輸入
- 可全篇下載目前正在閱讀的文章 (僅限部份 BBS 系統，已知 Maple 系統下可正常運作)
- 滑鼠選取 BBS 畫面上的文字，可右鍵選單自動以搜尋引擎搜尋，支援 OpenSearch plugins
- 支援自動回訊，正確設定可自動登入 BBS 站台
- 提供離座視窗上鎖功能，防偷窺
- 內建全球中文 BBS 站台列表，並且可隨時從網路更新
- 還有更多更多豐富功能...等待使用者慢慢發掘


PCMan Combo
連線 (C) 編輯 (E) 選項 (V) BBS最愛 (B) Web最愛 (A) 紀錄 (Y) ANSI編輯 (N) 工具 (T) 說明 (H)
新增連線分頁 (N)
開啟站台列表 (S)
中斷這個連線 (D)
關閉這個分頁 (C) Ctrl+Del
重新連線 (R) Alt+R
上次的連線 (P) Alt+F
開啟網頁 (O)
另存網頁 (A)
網頁列印設定 (P)
列印網頁 (P)
結束 (X)
站台列表
輸入關鍵字：(N)
尋找站台 (F)
新增站台 (A)
新增資料夾 (L)
設定 (E)
剪下 (T)
複製 (O)
貼上 (P)
刪除 (D)
加到我的最愛 (F)
加到首頁 (H)
更新列表 (U)
BBS 我的最愛
PCMan 相關站台
台大批踢踢實業坊 ptt.cc
KKCity bbs.kkcity.com.tw
BBS 站台列表
臺灣學術網 BBS
臺灣網際網 BBS
香港 BBS
大陸公眾網 BBS
大陸教育網 BBS
海外華語 BBS
2005新增位址
首頁
用滑鼠直接拖曳改變順序，按住 Ctrl 拖曳可以複製，右鍵選單尚有功能。要將某站台或資料夾移到其他資料夾內，只要拖曳到目標資料夾上停留一秒，資料夾就會自動展開
連線 (C)
關閉 (X)
站台設定
連線
站台選項
BBS自動回訊
主機位址：
bioinfo.nhri.org.tw
連接埠：
23
連線名稱：
CGC


站台列表
輸入關鍵字：(N)
尋找站台 (F)
新增站台 (A)
BBS 我的最愛
PCMan 相關站台
台大批踢踢實業坊 ptt.cc
GCG bioinfo.nhri.org.tw
BBS 站台列表
臺灣學術網 BBS
臺灣網際網 BBS
香港 BBS
大陸公眾網 BBS
大陸教育網 BBS
海外華語 BBS
2005新增位址
首頁
連線 (C)
關閉 (X)
GCG - PCMan Combo
連線 (C) 編輯 (E) 選項 (V) BBS最愛 (B) Web最愛 (A) 紀錄 (Y) ANSI編輯 (N) 工具 (T) 說明 (H)
telnet://bioinfo.nhri.org.tw
搜尋：
Orz
1.GCG
SunOS 5.9
login: petang
Password:
You have mail.
Welcome to GCG
Version 11.1.3-UNIX
Installed on solaris
Copyright (c) 1982 - 2006, Accelrys Inc.
All rights reserved.
Published research assisted by this software should cite:
GCG Version 11.1, Accelrys Inc., San Diego, CA
Databases available:
GenBank Release 165.0 (04/2008)
GenPept Release 165.0 (04/2008)
Refseq Release 28.0 (03/2008)
UniProt Release 13.2 (04/2008)
PROSITE Release 20.31 (04/2008)
Pfam Release 22.00 (06/2007)
Restriction Enzymes (REBASE) 804 (04/2008)
Technical support see: http://www.accelrys.com/support/
Online help: % genhelp or http://www.accelrys.com/support/bio/genhelp/
GCG System Support Environment Initialized.
v8803: petang [users/petang]>
連線時間：0 小時 1 分 28 秒
位址：telnet://bioinfo.nhri.org.tw

*https://filezilla-project.org/*



FileZilla The free FTP solution


Home

**FileZilla**
- Features
- Screenshots
- Download
- Documentation

**FileZilla Server**
- Download

**Community**
- Forum
- Project page
- Wiki

**General**
- FAQ
- Contact
- License
- Privacy Policy

**Development**
- Source code
- Nightly builds
- Translations
- Version history
- Changelog
- Issue tracker

**Other projects**
- libfilezilla
- Octochess

PayPal DONATE

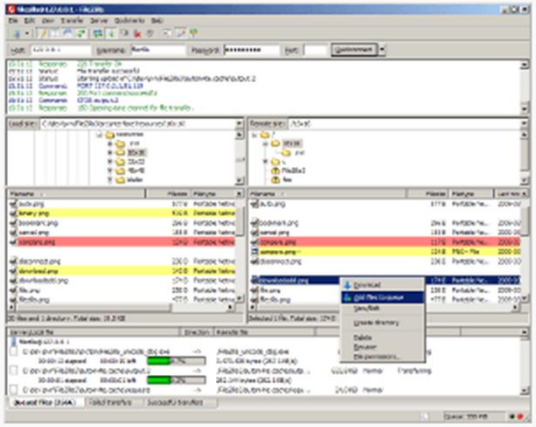
Sponsors:

NDC HOST


# Download FileZilla Client for Windows (64bit)

The latest stable version of FileZilla Client is 3.22.1

Please select the file appropriate for your platform below.



**Windows (64bit)**

Download FileZilla Client

This installer may include bundled offers. Check below for more options.

The 64bit versions of Windows Vista, 7, 8, 8.1 and 10 are supported.

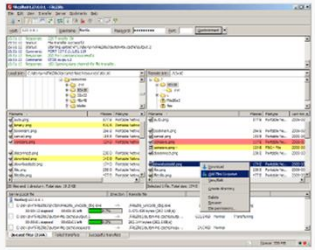
**More download options**

Other platforms:

Not what you are looking for?

Show additional download options

## Download FileZilla Client



The latest stable version of FileZilla Client is 3.22.1

Please select the file appropriate for your platform below.

**Windows (64bit)**

- FileZilla_3.22.1_win64-setup.exe (recommended)
- FileZilla_3.22.1_win64.zip

The 64bit versions of Windows Vista, 7, 8, 8.1 and 10 are supported.

**Windows (32bit)**

- FileZilla_3.22.1_win32-setup.exe (recommended)
- FileZilla_3.22.1_win32.zip

The 32bit versions of Windows Vista, 7, 8, 8.1 and 10 are supported.

**Mac OS X**

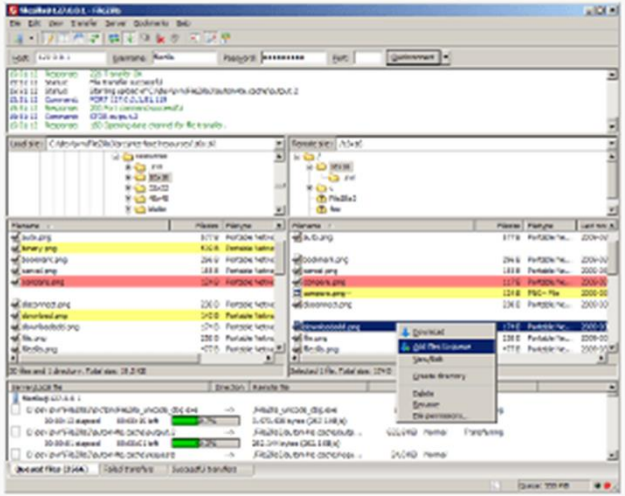
- FileZilla_3.22.1_macosx-x86.app.tar.bz2

Requires OS X 10.7 or newer

# Download FileZilla Client

The latest stable version of FileZilla Client is 3.22.1

Please select the file appropriate for your platform below.



### Windows (64bit)

FileZilla_3.22.1_win64-setup.exe (recommended)

FileZilla_3.22.1_win64.zip

The 64bit versions of Windows Vista, 7, 8, 8.1 and 10 are supported.

### Windows (32bit)

FileZilla_3.22.1_win32-setup.exe (recommended)

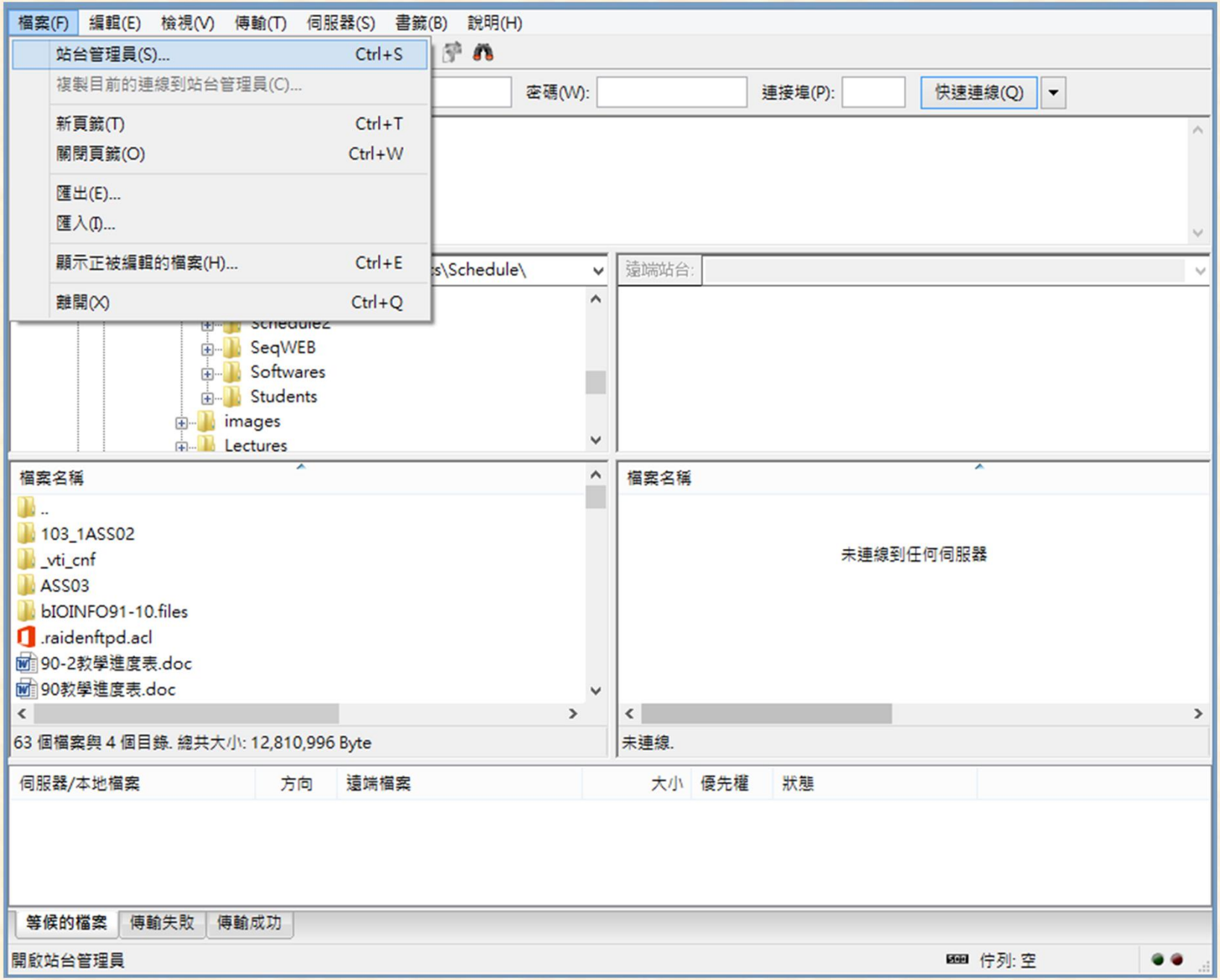
FileZilla_3.22.1_win32.zip

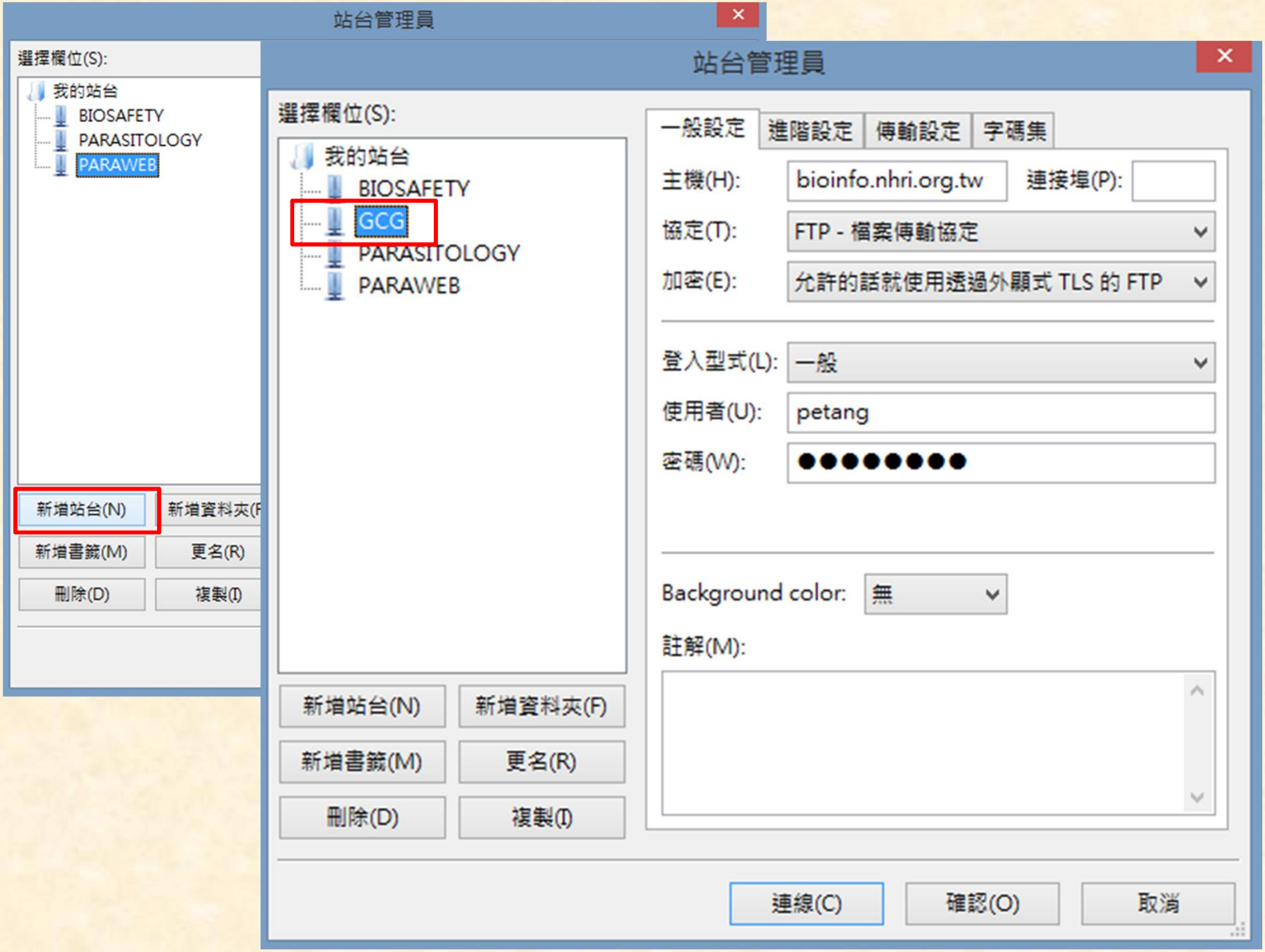
The 32bit versions of Windows Vista, 7, 8, 8.1 and 10 are supported.

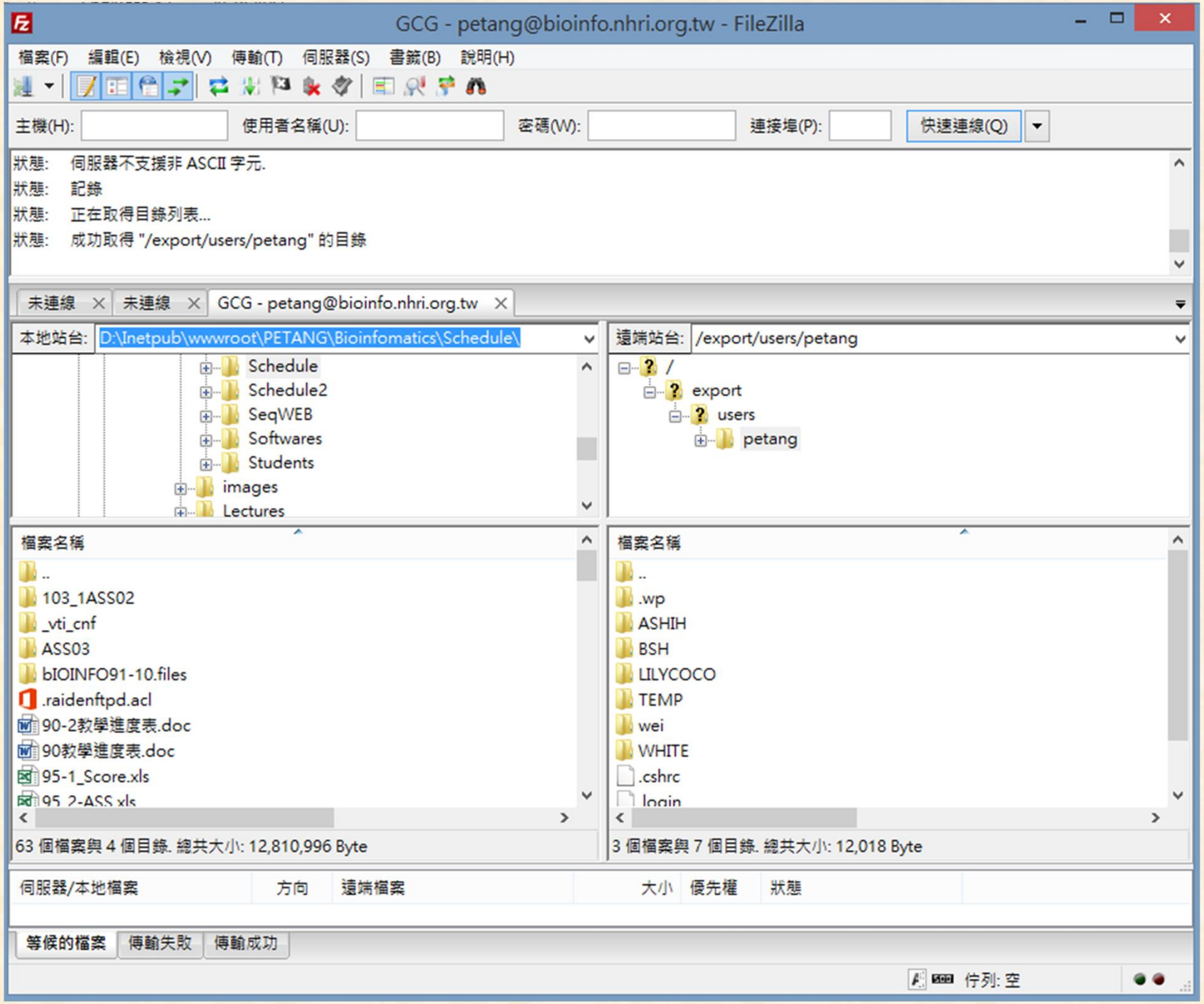
### Mac OS X

FileZilla_3.22.1_macosx-x86.app.tar.bz2

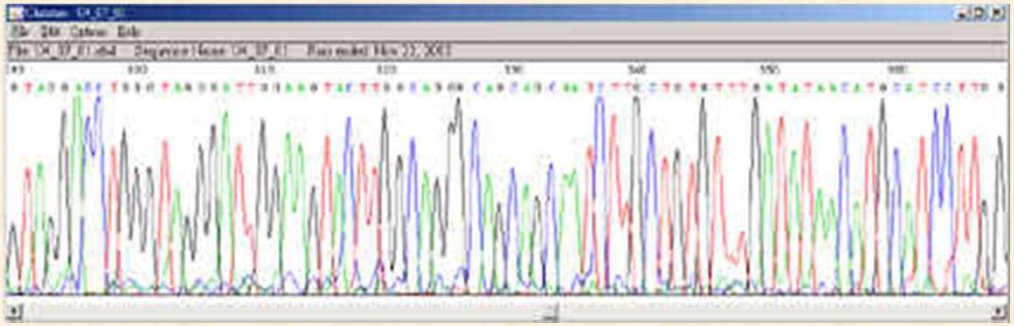
Requires OS X 10.7 or newer


檔案(F)
編輯(E)
檢視(V)
傳輸(T)
伺服器(S)
書籤(B)
說明(H)
站台管理員(S)... Ctrl+S
複製目前的連線到站台管理員(C)...
新頁籤(T) Ctrl+T
關閉頁籤(O) Ctrl+W
匯出(E)...
匯入(I)...
顯示正被編輯的檔案(H)... Ctrl+E
離開(X) Ctrl+Q
密碼(W):
連接埠(P):
快速連線(Q)
遠端站台:
SeqWEB
Softwares
Students
images
Lectures
檔案名稱
103_1ASS02
_vti_cnf
ASS03
bIOINFO91-10.files
.raidenftpd.acl
90-2教學進度表.doc
90教學進度表.doc
63 個檔案與 4 個目錄. 總共大小: 12,810,996 Byte
未連線到任何伺服器
未連線.
伺服器/本地檔案
方向
遠端檔案
大小
優先權
狀態
等候的檔案
傳輸失敗
傳輸成功
開啟站台管理員
佇列: 空


站台管理員
選擇欄位(S):
我的站台
BIOSAFETY
PARASITOLOGY
PARAWEB
新增站台(N)
新增資料夾(F
新增書籤(M)
更名(R)
刪除(D)
複製(I)
站台管理員
選擇欄位(S):
我的站台
BIOSAFETY
GCG
PARASITOLOGY
PARAWEB
新增站台(N)
新增資料夾(F)
新增書籤(M)
更名(R)
刪除(D)
複製(I)
一般設定
進階設定
傳輸設定
字碼集
主機(H):
bioinfo.nhri.org.tw
連接埠(P):
協定(T):
FTP - 檔案傳輸協定
加密(E):
允許的話就使用透過外顯式 TLS 的 FTP
登入型式(L):
一般
使用者(U):
petang
密碼(W):
Background color:
無
註解(M):
連線(C)
確認(O)
取消


GCG - petang@bioinfo.nhri.org.tw - FileZilla
檔案(F) 編輯(E) 檢視(V) 傳輸(T) 伺服器(S) 書籤(B) 說明(H)
主機(H): 使用者名稱(U): 密碼(W): 連接埠(P): 快速連線(Q)
狀態: 伺服器不支援非 ASCII 字元.
狀態: 記錄
狀態: 正在取得目錄列表...
狀態: 成功取得 "/export/users/petang" 的目錄
未連線
未連線
GCG - petang@bioinfo.nhri.org.tw
本地站台: D:\Inetpub\wwwroot\PETANG\Bioinfomatics\Schedule\
Schedule
Schedule2
SeqWEB
Softwares
Students
images
Lectures
遠端站台: /export/users/petang
/
export
users
petang
檔案名稱
..
103_1ASS02
_vti_cnf
ASS03
bIOINFO91-10.files
.raidenftpd.acl
90-2教學進度表.doc
90教學進度表.doc
95-1_Score.xls
63 個檔案與 4 個目錄. 總共大小: 12,810,996 Byte
檔案名稱
..
.wp
ASHIH
BSH
LILYCOCO
TEMP
wei
WHITE
.cshrc
login
3 個檔案與 7 個目錄. 總共大小: 12,018 Byte
伺服器/本地檔案
方向
遠端檔案
大小
優先權
狀態
等候的檔案
傳輸失敗
傳輸成功
佇列: 空

# Softwares & Sequence Formats



| Program | Formats Default | Accept | Multiple sequence |
|---|---|---|---|
| **WWW** | text file | paste & Copy | |
| **SeqWEB** | text file | paste & copy | |
| **GCG** | ● GCG file | FASTA ●<br>GenBANK ● ●<br>EMBL ● ●<br>Staden<br>SwissProt | Multiple sequence file (msf)<br>Rich sequence file (rsf)<br>List files (lst) |
| **VectorNTI** | | | |
| **CLC Genomics** | | | |

# Importing and Exporting

You need a FTP program to transfer files between your PC and GCG.
The sequence file must be in “plain text” format.

**chopup:** converts a non-GCG format sequence file containing lines longer than 511 characters and as long as 32,000 characterters into a new file containing no longer than 50 characters.

**breakup:** reads a non-GCG format sequence file containing more than 350,000 sequence characterters and writes it as a set of separate, shorter, overlapping sequence files than can be analyzed by GCG.

**reformat:** rewrites sequence files, scoring matrix files, or enzyme data files so than they can be read by GCG programs.

**fromfasta:** reformats one or more sequences from FastA format into single sequence files in GCG format.

# Exercise 02-2

(A) Download naq.txt and psq.txt
(B) Transfer sequence files from your PC to GCG
(C) Chopup the sequence
(D) Reformat the sequence
(E) Edit the sequence

Create a folder “BIO” in your hard disk
Start WsFTP (ftp://bioinfo.nhri.org.tw)
Upload “naq.txt” & “psq.txt” to GCG
Start Netterm
Start GCG
Chopup “naq.txt” & “psq.txt”
Reformat “naq.dat” or “psq.dat”
Cat “naq.txt” or “psq.txt”

# Exercise 02-3

## Sequence Manipulation in GCG UNIX

Use the database searching techniques you learned today to retrieve the reference sequence

**Homo sapiens LEGUMAIN**

and the amino acid sequence of

**ALL LEGUMAIN**

From NCBI
And then transfer the sequence(s) to
GCG Unix (in GCG format)

There are many different ways to DO it.
You can have your lunch now if you can make it.

# ASSIGNMENT 1.

Use the Entrez searching techniques you learned today to retrieve the
**Reference sequence and**
**the corresponding amino acid sequences of**

**All the subclasses of *Homo sapiens* cyclophilin**

Transfer the sequences to GCG Unix,
Transform the sequences to GCG format

E-mail
1. The steps (including URL of WWW sites) you used and
2. The sequences in GCG format as attached file to
***petang@mail.cgu.edu.tw*** before next Thursday 1200
****郵件主旨： Bioinfo2016ASS1 – (學號)