EGTDC SOFTWARE DOCUMENTATION PAGES

Software Documentation Home | Google this title | Google this category | PermaLink | Edit | Add a Comment (0)

BIO-LINUX FAQ (V2.0)

Frequently Asked Questions about getting and using EGTDC's Bio-Linux.

EGTDC Bio-Linux FAQ v2.0 (February 2003)

The Bio-Linux FAQ is maintained by **Dan Swan**

The Bio-Linux project is part of the EGTDC and aims to provide all PI and co-PI's with EG awards a dedicated bioinformatics workstation running Linux. The aim is to provide a core set of applications and SOPs for installation of other software along with sample data, test suites and extensive documentation.

The following people have generously contributed to the FAQ:

Nic Bertrand, Joe Wood, Milo Thurston, Dawn Field, Bill Tyne, Rob Griffiths, Kerr Wall

Introduction and Installation of Bio-Linux

Post install support, and questions

Techinical details and advanced information

Credits and additional documentation

1. **What is Bio-Linux?**

   Bio-Linux is a modified distribution of RedHat Linux which aims to make Linux easily deployed in research labs as bioinformatics workstations. It is set up to be user friendly to a casual user, yet powerful enough to deal with intensive, expert use.

2. **How can Bio-Linux help me?**

   Bio-Linux provides you with a dedicated bioinformatics research platform, pre-loaded with many applications you will use in the course of your research. By centralising these packages and providing you with dedicated, powerful hardware on which to run them the speed and scope of your research can be improved.

3. **How do I know if I need Bio-Linux?**

   Are you fed up with slow, web-based interfaces to bioinformatics tools? Do you see Linux being mentioned frequently and wish you could test it out? Are you frustrated by the inability to run research software under Windows or MacOS? Are you annoyed by the instability of software on your current system? Are you looking to get into the fastest growing computer movement on the planet?

   If you answered yes to any of these questions Bio-Linux is for you!

4. **What Bioinformatics software is on this machine?**

   Open Source software included:

   ACT - Sanger Centre Genome Comparison tool
   Artemis - Sanger Center Genome Annotation Tool
   ARB - 16S phylogeny analysis package
   NCBI Blast - Standalone blast from NCBI
   Netblast - Blast client from NCBI for doing remote blast
   FASTA - Standalone fasta
   ClustalX/clustalw - Mutliple Alignment Tool
   T-Coffee/Seaview - Mutliple Alignment Tool
   EMBOSS - Package of sequence analysis tools
   maxd - transcriptomics software and database
   Jalview - Mulitple Alignment Editor
   Phylip - Phylogenetic Analysis Package
   Treetool - View phylogenetic trees
   Bioperl - Perl Development code for bioinformatics
   TRIBE-MCL - Clustering of protein sequences
   MSPcrunch/Blixem - Blast report post-processor/Viewer
   Dotter - Generate and View Dotplots
   Readseq - File Format converter
   Rasmol - Protein Structure Viewer

   Free software that requires an academic licence:

   phred/phrap cross_match/swat/consed - Trace file analysis and sequence assembly
   CAP - EST clustering

   Commercial software:

   Genespring (Silicon Genetics) - transcriptomics software

5. **Is Bio-Linux free?**

   Yes. Bio-Linux relies heavily on "Open Source" software, and as such is free to redistribute.

6. **What kind of hardware do I need to run Bio-Linux?**

   Bio-Linux is currently hardware dependant, therefore we very strongly suggest that you purchase the same hardware that Bio-Linux was developed on for optimal compatibility. Bio-Linux has been developed on Dell Optiplex GX240/GX260 machines with very specific hardware, but it is important that you email **us** so that we may advise you on the precise hardware purchase.

It may be possible to install Bio-Linux on other hardware configurations but you will need to have a competent Linux adminsitrator on-hand for post install configuration. We cannot support updates of Bio-Linux, nor provide technical troubleshooting for non-standard hardware.

7. **How do I get Bio-Linux?**

   Please contact the EGTDC **helpdesk** for more information. Each PI and co-PI from EG funded labs is eligible for a Bio-Linux computer provided through the Programme (Distributed in early 2003).
   The application form for a Bio-Linux distribution can be found **here**

   Pre installation information can be found **here**

   Once the forms have been completed a CD-ROM and floppy disk containing the install will be sent to you.

8. **Why must Bio-Linux be downloaded from the EGTDC instead of placed on an install CD-ROM?**

   Most distributions of Linux are available off CD-ROM, but Bio-Linux is not, at least not as of yet. Bio-Linux has been created using the **SystemImager** application. As such, Bio-Linux is not a Linux 'distribution' but a saved 'image' of a complete installation of Linux held on a specific computer at the EGTDC. While SystemImager is an invaluable tool for creating 'clones' of a computer, it currently does not support placing this image onto a disk for subsequent installation (it would require a DVD to place the entire distribution of a disk). The only way to 'get' Bio-Linux is to buy a machine capable of running this image and installing the image. In the future, we hope that Bio-Linux will evolve into a distribution in the full sense of the word.

   It is important to note that the install CD does not contain any part of the Bio-Linux distribution it is merely a bootable image that allows you to connect to the distribution server and obtain the Bio-Linux system image.

9. **What is all the network information on the Bio-Linux application? Why do you need this?**

   On your Bio-Linux application form you will be asked for the following information:

   - Hostname of the new machine - this is required for correct configuration of your clone. This **must** match the "real" name of the machine, and not some arbitrary name you have decided to call it. Your CS people will know this information
   - Domain name - again this is required for correct clone configuration, it will probably match the information after the @ on your email address, but may contain extra subdomains. Your CS people will know this information.
   - Device - this refers to the network card in your machine. Generally this will be eth0 (ethernet device 0).
   - IP of new machine - The Internet Protocol address of the machine. We need this information in order to open our firewall and allow your install to proceed.
   - Netmask of new machine - most likely to be 255.255.255.0 but ask your CS people. We need this information for post install configuration.
   - Network IP for new machine - this will most likely be of the form XXX.XXX.XXX.0
   - Broadcast IP for new machine - this will most likely be of the form XXX.XXX.XXX.255
   - Gateway IP for new machine - please ask your CS staff for this
   - Primary DNS server IP - the Domain Name Server for your organisation
   - Secondary DNS server - the secondary Domain Name Server for your organisation

   It is essential that this information comes from your CS department and is not an 'educated guess' as to what you think those values might be. This is **especially** important for the host-name.

   All the information is an absolute requirement of the SystemImager install process. We keep a copy of this information for our records, in the strictest confidence. No-one outside the Bio-Linux project will see this information.

10. **Now I have my floppy and CDROM, how do I install?**

    Full installation instructions can be found **here**

11. **How do I decide which user group I should be in? What level of support does each user group entail?**

    We use a categorisation system with Bio-Linux users in order to understand what levels of service and support we need to offer each site.

- User Group 1: You fall into this group if you have no Linux experience and cannot find Linux support from your own IT department. Your only concern with Bio-Linux is to have the tools on your dekstop. You are not interested in systems administrations in the task. You do not envisage anything more than infrequent updates.

  Your Bio-Linux support will be transparent and complete. There will be no need to worry about updates, and we will not ask too much of you unless a major update is required and we need someone on-hand to babysit an upgrade.

- User Group 2: You fall into this group if you have no, or some Linux experience, but there is someone on hand to help you out if you get stuck. You want to use the tools on Bio-Linux but wouldn't mind learning some basic systems administration as well. You would like to keep your Bio-Linux desktop up to date.

  Your Bio-Linux support will be tailored to your needs. We will instruct you on how best to care for your system, and will always be on hand to deal with any questions you may have. We understand that you are interested in taking things further, and will support you in this.

- User Group 3: You have some Linux experience, and generally don't need too much help from others to fix problems, as you know where to look online. You want to use the tools on Bio-Linux but are also willing to feed back to the Bio-Linux list about features you would like to see added. You are happy to do systems administration and would like to customise Bio-Linux to suit your specific needs.

  Your Bio-Linux support will be tailored to your needs. We anticipate that you will not require our support too much. We will keep you appraised of updates, and allow you the power to control their implementation. We will involve you in future development roadmap discussions and you will contribute to the next versions of Bio-Linux with your suggestions.

12. **What if I'm not funded by the EG Programme but I want to have Bio-Linux?**

    If you fall into User Group 3 (Advanced), we would love to hear from you about potential use/development of Bio-Linux. If you fall into User Groups 1 or 2, the EGTDC is currently devising mechanisms by which it can provide services to NERC and non-NERC awardees either under a financial model or through shared posts. Please feel free to contact us for more information.

13. **Where can I get more help if I need it?**

    To get more help you can join the envgen-informatics mailing list. You can subscribe to the mailing list **here**

14. **What other common tools are included?**

    Bio-Linux includes the common Linux/UNIX tools that you expect.

    Editors: pico, vi, vim, gvim, emacs, kedit, kate, kwrite, gedit
    Databases: MySQL, postgres
    Programming Languages: Perl, PHP, python, Java, compilers for C, C++
    Web browsers: Mozilla, Galleon, Konqueor (KDE), Netscape
    Window managers: Ximian Gnome, KDE (Gnome is customised for Bio-Linux)

15. **How can I suggest improvements?**

    Either post comments to the **Bio-Linux mailing list** or post comments onto the documents on this web site.

16. **How can I contribute to Bio-Linux?**

    Contribution to Bio-Linux is strongly encouraged. If you have any feedback at all whether it is on the installation process, software packages included or general comments, please feed back to the **mailing list**.
    If you have RPM versions of bioinformatics software, or you have released a program that you would like to see included, please get in touch.
    If you have experience in building distributions of Linux we would be extremely interested in hearing from you.

17. **How will Bio-Linux deal with distribution upgrades of Linux?**

    Ximian Red Carpet will keep your Bio-Linux workstation up to date for patches, bug fixes, security fixes and package updates, but it will not upgrade major releases of the distribution itself (for instance it will not migrate you from RedHat 7.3 to 8.0). When a major distribution upgrade occurs the EGTDC will implement it on the image server and then alert Bio-Linux adminitrators to re-image the machine. This will NOT affect data stored in /home or /db so your data is safe, although we advise you strongly to back up your data before upgrading. In the future when Bio-Linux becomes a distribution in its own right we will be able to send out upgrades via CD-ROM.

18. **Why was RedHat chosen and which version is it based on?**

    RedHat is a mature distribution of Linux and Red Hat Linux has a 52% marketshare among Linux operating systems, according to Deutsche Bank Alex. Brown Estimates and IDC. The Bio-Linux developers have many years experience building and maintaining RedHat systems as workstations and servers and it seemed natural to extend this in Bio-Linux.

    Currently Bio-Linux 2.0 is based on RedHat 7.3. Future releases of Bio-Linux will be based on versions of Linux software current at the release time.

19. **Is there a list of rpm's installed on the system?**

    RPM stands for 'RedHat Package Manager' and is a simple method of supplying software to users without the need for specialised knowledge for compiling utilities. The full output of 'rpm -qa' can be found **here**

20. **May I have a dual boot system? I want to run windows applications on the workstation**

    We understand that it is likely that there are some Windows applications you may wish to run on the Bio-Linux workstations, however we cannot provide a mechanism for dual boot machines due to technical reasons. Whilst it might be possible to create a dual boot machine, you would ulitimately be responsible for it - including the dual boot setup.

    If you are interested in running Windows software we recommend installing **VMware** which emulates a full Windows system inside your Bio-Linux workstation and would allow your existing applications to run.

    If its Office functionality you are after then OpenOffice will probably meet your needs as it includes a word processor, spreadsheet and presentation packages, as well as other features. It is compatible with the Microsoft Office Suite. If you absolutely must run Office then **Crossover Office** allows Bio-Linux to run the Microsoft Office Suite. VMware and Crossover Office are not included in the Bio-Linux distribution but we will support the installation of VMware if a group requires it.

21. **Can I install Bio-linux on whatever machine I please?**

    Yes, you are more than welcome to try, but don't expect it to work perfectly. If you choose to do this, then you should note the following:

    - The Bio-Linux install CD may not support your video card, and the install may not work.
    - You may see error messages instead of getting the beeping noises you may be expecting after a successful install. The install may have worked anyway, though.
    - When you reboot, a program called kudzu will start up. It will point out that your hardware has changed, and ask if you want to remove some configurations and configure some new hardware. You MUST remove old configurations and configure the new hardware, or you will find yourself in trouble. Not to worry, though - it's fun and easy!
    - Once kudzu is finished, you will be able to run the setup.perl script as usual. However, you must NOT run /sbin/telinit 5 yet. Before you do, run xfree86 -configure. If X still does not work then seek further help from your local IT support department or on the internet (e.g. www.xfree86.org). The details of configuring X are beyond the scope of this document.
    - Once you're up and running we will not be able to support you, or offer you updates, you will literally be responsible for every aspect of the systems adminstration of that machine.

22. **Why does GRUB have a password?**

    We are particularly keen on security in Bio-Linux. Each machine comes with a custom firewall, and as few services running as we can get away with. Whilst most people consider security at the level of the network, very few consider security at the site of installation. GRUB is password protected to stop people from booting into 'single user' mode, which would effectively allow them full control of your machine without a password. We consider this of sufficient concern to install GRUB with a password.

23. **What if I don't like the way the machine is set up, and don't approve of your policies?**

    You are more than welcome to change whatever you like, but when you do, we won't be able to provide you with support. This is not because we don't want to, but because the difficulty of supporting somany different configurations is too much for our small number of staff. We also can't guarantee that your changes will be persistant through future upgrades.

24. **What services are running, and how secure is the machine?**

    By default sshd is running and allows connections. SMTP is also running but is not accessible externally and cannot be used as an open relay (ie for spamming). MySQL is running but firewalled from the outside world.
    The machine does not respond to pings or traceroute requests, and using various nmap options to scan the machine will only reveal port 22 (sshd) to be open. Apache is installed but not enabled by default. We believe the machine is secure from computer hackers providing OpenSSH is kept up to date.

25. **What window manager is Ximian?**

    Ximian GNOME (aka Ximian Desktop) is used on Bio-Linux to provide a user-friendly desktop backed up with a comprehensive email client/PIM (Ximian Evolution) and allow Bio-Linux workstations to be easily update via Ximian Red Carpet. The Ximian extensions sit on top of Gnome 1.4.

26. **Credits: Who did all this work?**

    So far, this is part of the EGTDC implementation plan in the Field lab at CEH Oxford. Milo Thurston has developed the 'gold server' with SystemImager and the technical work to create the Bio-Linux image. Many have worked as beta testers and have contributed by suggesting new tools and features to add. This includes researchers inside and outside CEH Oxford. Bio-Linux 2.0 development was taken up by Dr Dan Swan in February 2003.

27. **Related Documents:**

    - **Bioinformatics software in Bio-Linux**
    - **Requesting Bio-Linux**
    - **Install Instructions**
    - **Additional install instructions once you are registered for a download:**
    - **Information you should know about Linux and how to use Bio-Linux**
    - **More tips on using Bio-Linux**