Sequence analysis and bioinformatics using Debian GNU/Linux

Andreas Tille

Libre Software Meeting

LSM, Nantes 2009
Overview

1. Debian Med
   - Debian Pure Blend for medical care and health science
   - Why Debian

2. Implementation
   - Available packages
   - Biological databases

3. Looking beyond
   - Alternatives and prospectus
1 Debian Med
   - Debian Pure Blend for medical care and health science
   - Why Debian

2 Implementation
   - Available packages
   - Biological databases

3 Looking beyond
   - Alternatives and prospectus
Overview

1. Debian Med
   - Debian Pure Blend for medical care and health science
   - Why Debian

2. Implementation
   - Available packages
   - Biological databases

3. Looking beyond
   - Alternatives and prospectus
1. Debian Med
   - Debian Pure Blend for medical care and health science
   - Why Debian

2. Implementation
   - Available packages
   - Biological databases

3. Looking beyond
   - Alternatives and prospectus
Scope of Debian Med

- **Free management systems for patients in medical practice and hospitals are rare**
  - **GNUmed**: Patient record documentation for general practitioners
  - **MedinTux**: Practice management system written for French health care system
  - **Vista**: Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
  - **Care2x**: Web based hospital management system
  - **Others**...

- However, people who hear the sound “Debian Med” just assume we provide a practice management system …
- … even if you tell them explicitly it is not
- So what are the real strengths of Debian Med?
Scope of Debian Med

- Free management systems for patients in medical practice and hospitals are rare
  - GNUmed: Patient record documentation for general practitioners
  - MedinTux: Practice management system written for French health care system
  - Vista: Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
  - Care2x: Web based hospital management system
  - Others . . .

However, people who hear the sound “Debian Med” just assume we provide a practice management system . . .

. . . even if you tell them explicitly it is not

So what are the real strengths of Debian Med?
Scope of Debian Med

- Free management systems for patients in medical practice and hospitals are rare
  - GNUmed: Patient record documentation for general practitioners
  - MedinTux: Practice management system written for French health care system
  - Vista: Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
  - Care2x: Web based hospital management system
  - Others . . .

- However, people who hear the sound “Debian Med” just assume we provide a practice management system . . .
- . . . even if you tell them explicitly it is not
- So what are the real strengths of Debian Med?
Scope of Debian Med

- Free management systems for patients in medical practice and hospitals are rare
  - GNUmed: Patient record documentation for general practitioners
  - MedinTux: Practice management system written for French health care system
  - Vista: Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
  - Care2x: Web based hospital management system
  - Others . . .

However, people who hear the sound “Debian Med” just assume we provide a practice management system . . .

. . . even if you tell them explicitly it is not

So what are the real strengths of Debian Med?
Scope of Debian Med

- Free management systems for patients in medical practice and hospitals are rare
  - **GNUmed** Patient record documentation for general practitioners
  - **MedinTux** Practice management system written for French health care system
  - **Vista** Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
  - **Care2x** Web based hospital management system
  - Others . . .

However, people who hear the sound “Debian Med” just assume we provide a practice management system . . .

. . . even if you tell them explicitly it is not

So what are the real strengths of Debian Med?
Scope of Debian Med

- Free management systems for patients in medical practice and hospitals are rare
  - GNUmed: Patient record documentation for general practitioners
  - MedinTux: Practice management system written for French health care system
  - Vista: Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
  - Care2x: Web based hospital management system
  - Others . . .

- However, people who hear the sound “Debian Med” just assume we provide a practice management system . . .
- . . . even if you tell them explicitly it is not
- So what are the real strengths of Debian Med?
Scope of Debian Med

- Free management systems for patients in medical practice and hospitals are rare
  - GNUmed: Patient record documentation for general practitioners
  - MedinTux: Practice management system written for French health care system
  - Vista: Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
  - Care2x: Web based hospital management system
  - Others...

- However, people who hear the sound “Debian Med” just assume we provide a practice management system...
- ...even if you tell them explicitly it is not
- So what are the real strengths of Debian Med?
Free management systems for patients in medical practice and hospitals are rare

- GNUmed: Patient record documentation for general practitioners
- MedinTux: Practice management system written for French health care system
- Vista: Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
- Care2x: Web based hospital management system
- Others . . .

However, people who hear the sound “Debian Med” just assume we provide a practice management system . . .

. . . even if you tell them explicitly it is not

So what are the real strengths of Debian Med?
Scope of Debian Med

- Free management systems for patients in medical practice and hospitals are rare
  - GNUmed: Patient record documentation for general practitioners
  - MedinTux: Practice management system written for French health care system
  - Vista: Comprehensive software suite for hospitals (U.S. Department of Veterans Affairs)
  - Care2x: Web based hospital management system
  - Others . . .

- However, people who hear the sound “Debian Med” just assume we provide a practice management system . . .
- . . . even if you tell them explicitly it is not
- So what are the real strengths of Debian Med?
Medical imaging

- Debian Med can only include *existing* software
- Fair amount of high quality Free Software for medical imaging
  - Aeskulap, Amide: Medical image viewers
  - Dcmtk: OFFIS DICOM toolkit
  - Sofa: Simulation Open Framework Architecture
  - Fsl: analysis tools for brain imaging
  - ...

- Complete overview on [Debian Med imaging tasks page](#)
Medical imaging

- Debian Med can only include *existing* software
- Fair amount of high quality Free Software for medical imaging
  - Aeskulap, Amide: Medical image viewers
  - Dcmtk: OFFIS DICOM toolkit
  - Sofa: Simulation Open Framework Architecture
  - Fsl: analysis tools for brain imaging
  - ...
- Complete overview on *Debian Med imaging tasks page*
Debian Med can only include *existing* software

Fair amount of high quality Free Software for medical imaging

- Aeskulap, Amide: Medical image viewers
- Dcmtk: OFFIS DICOM toolkit
- Sofa: Simulation Open Framework Architecture
- Fsl: analysis tools for brain imaging
- ... 

Complete overview on *Debian Med imaging tasks page*
Debian Med can only include *existing* software

Fair amount of high quality Free Software for medical imaging

- Aeskulap, Amide: Medical image viewers
- Dcmtk: OFFIS DICOM toolkit
- Sofa: Simulation Open Framework Architecture
- Fsl: analysis tools for brain imaging
  ...

Complete overview on *Debian Med imaging tasks page*
Debian Med can only include *existing* software

Fair amount of high quality Free Software for medical imaging

- Aeskulap, Amide: Medical image viewers
- Dcmtk: OFFIS DICOM toolkit
- Sofa: Simulation Open Framework Architecture
- Fsl: analysis tools for brain imaging
- ...

Complete overview on [Debian Med imaging tasks page](#)
Debian Med can only include *existing* software

- Fair amount of high quality Free Software for medical imaging
  - Aeskulap, Amide: Medical image viewers
  - Dcmtk: OFFIS DICOM toolkit
  - Sofa: Simulation Open Framework Architecture
  - Fsl: analysis tools for brain imaging
  - ...

- Complete overview on *Debian Med imaging tasks page*
Debian Med can only include *existing* software

Fair amount of high quality Free Software for medical imaging
- Aeskulap, Amide: Medical image viewers
- Dcmrk: OFFIS DICOM toolkit
- Sofa: Simulation Open Framework Architecture
- Fsl: analysis tools for brain imaging

... 

Complete overview on [Debian Med imaging tasks page](#)
Debian Med can only include *existing* software

Fair amount of high quality Free Software for medical imaging

- Aeskulap, Amide: Medical image viewers
- Dcmtk: OFFIS DICOM toolkit
- Sofa: Simulation Open Framework Architecture
- Fsl: analysis tools for brain imaging
- ...

Complete overview on [Debian Med imaging tasks page](#)
Molecular and structural biology, bioinformatics

- **Most established branch of Debian Med** because of good coverage by upstream software
  - **Fostering**
    - Development at universities
    - Organised funding
  - **Hindering**
    - Advertising for proprietary software
    - Different preferences of initiators

Andreas Tille
(Libre Software Meeting)
Sequence analysis and bioinformatics
LSM, Nantes 2009  6 / 32
Molecular and structural biology, bioinformatics

- Most established branch of Debian Med because of good coverage by upstream software
- Fostering
  - Development at universities
  - Organised funding
- Hindering
  - Advertising for proprietary software
  - Different preferences of initiators
Molecular and structural biology, bioinformatics

- Most established branch of Debian Med because of good coverage by upstream software
- Fostering
  - Development at universities
  - Organised funding
- Hindering
  - Advertising for proprietary software
  - Different preferences of initiators
Molecular and structural biology, bioinformatics

- Most established branch of Debian Med because of good coverage by upstream software
- Fostering
  - Development at universities
  - Organised funding
- Hindering
  - Advertising for proprietary software
  - Different preferences of initiators

Andreas Tille (Libre Software Meeting)
Molecular and structural biology, bioinformatics

- Most established branch of Debian Med because of good coverage by upstream software
- Fostering
  - Development at universities
  - Organised funding
- Hindering
  - Advertising for proprietary software
  - Different preferences of initiators
Molecular and structural biology, bioinformatics

- Most established branch of Debian Med because of good coverage by upstream software
- Fostering
  - Development at universities
  - Organised funding
- Hindering
  - Advertising for proprietary software
  - Different preferences of initiators
Molecular and structural biology, bioinformatics

- Most established branch of Debian Med because of good coverage by upstream software
- Fostering
  - Development at universities
  - Organised funding
- Hindering
  - Advertising for proprietary software
  - Different preferences of initiators
Selected metapackages of Debian Med

Number of dependencies of selected Debian Med metapackages

- Microbiology
- Imaging
- Practice

Andreas Tille (Libre Software Meeting) Sequence analysis and bioinformatics LSM, Nantes 2009
Top 10 posters on debian-med@lists.debian.org

Andreas Tille (Libre Software Meeting) | Sequence analysis and bioinformatics | LSM, Nantes 2009
1. Debian Med
   - Debian Pure Blend for medical care and health science
   - Why Debian

2. Implementation
   - Available packages
   - Biological databases

3. Looking beyond
   - Alternatives and prospectus
Differences to commercial distributions

Commercial distributor

Debian
Differences to commercial distributions

Commercial distributor

Company

Structure

Debian

Organisation

Certified Oracle, SAP, etc. Runs in principle

Beginners Preferred by Administrators

Rpm Packages Deb

CDs, Service Sells Nothing

Business plan Release If 0 RC-bugs

Employees Persons Volunteers

Company Structure Organisation
## Differences to commercial distributions

<table>
<thead>
<tr>
<th></th>
<th>Commercial distributor</th>
<th>Debian</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Company</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Employees</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Structure</strong></td>
<td>Structure</td>
<td>Organisation</td>
</tr>
<tr>
<td><strong>Persons</strong></td>
<td>Persons</td>
<td>Volunteers</td>
</tr>
</tbody>
</table>
## Differences to commercial distributions

<table>
<thead>
<tr>
<th>Company</th>
<th>Structure</th>
<th>Organisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Employees</td>
<td>Persons</td>
<td>Volunteers</td>
</tr>
<tr>
<td>CDs, Service</td>
<td>Sells</td>
<td>Nothing</td>
</tr>
</tbody>
</table>

- **Commercial distributor**
  - Company
  - Employees
  - CDs, Service

- **Debian**
  - Structure
  - Persons
  - Sells
  - Nothing
Differences to commercial distributions

<table>
<thead>
<tr>
<th>Commercial distributor</th>
<th>Debian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>Structure</td>
</tr>
<tr>
<td>Employees</td>
<td>Persons</td>
</tr>
<tr>
<td>CDs, Service</td>
<td>Sells</td>
</tr>
<tr>
<td>Business plan</td>
<td>Release</td>
</tr>
<tr>
<td></td>
<td>Organisation</td>
</tr>
<tr>
<td></td>
<td>Volunteers</td>
</tr>
<tr>
<td></td>
<td>Nothing</td>
</tr>
<tr>
<td></td>
<td>If 0 RC-bugs</td>
</tr>
</tbody>
</table>
## Differences to commercial distributions

<table>
<thead>
<tr>
<th>Commercial distributor</th>
<th>Debian</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Company</strong></td>
<td>Structure</td>
</tr>
<tr>
<td><strong>Employees</strong></td>
<td>Persons</td>
</tr>
<tr>
<td><strong>CDs, Service</strong></td>
<td>Sells</td>
</tr>
<tr>
<td><strong>Business plan</strong></td>
<td>Release</td>
</tr>
<tr>
<td><strong>Certified</strong></td>
<td>Oracle, SAP, etc.</td>
</tr>
</tbody>
</table>

*Andreas Tille* (Libre Software Meeting)  
Sequence analysis and bioinformatics  
LSM, Nantes 2009
## Differences to commercial distributions

<table>
<thead>
<tr>
<th>Commercial distributor</th>
<th>Debian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>Structure</td>
</tr>
<tr>
<td>Employees</td>
<td>Persons</td>
</tr>
<tr>
<td>CDs, Service</td>
<td>Sells</td>
</tr>
<tr>
<td>Business plan</td>
<td>Release</td>
</tr>
<tr>
<td>Certified</td>
<td>Oracle, SAP, etc.</td>
</tr>
<tr>
<td>Beginners</td>
<td>Preferred by</td>
</tr>
</tbody>
</table>
### Differences to commercial distributions

<table>
<thead>
<tr>
<th>Commercial distributor</th>
<th>Debian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>Structure</td>
</tr>
<tr>
<td>Employees</td>
<td>Persons</td>
</tr>
<tr>
<td>CDs, Service</td>
<td>Sells</td>
</tr>
<tr>
<td>Business plan</td>
<td>Release</td>
</tr>
<tr>
<td>Certified</td>
<td>Oracle, SAP, etc.</td>
</tr>
<tr>
<td>Beginners</td>
<td>Preferred by</td>
</tr>
<tr>
<td>Rpm</td>
<td>Packages</td>
</tr>
<tr>
<td>Organisation</td>
<td>Volunteers</td>
</tr>
<tr>
<td></td>
<td>Nothing</td>
</tr>
<tr>
<td></td>
<td>If 0 RC-bugs</td>
</tr>
<tr>
<td></td>
<td>Runs in principle</td>
</tr>
<tr>
<td></td>
<td>Administrators</td>
</tr>
<tr>
<td></td>
<td>Deb</td>
</tr>
</tbody>
</table>
Differences to commercial distributions

<table>
<thead>
<tr>
<th>Commercial distributor</th>
<th>Debian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>Structure</td>
</tr>
<tr>
<td>Employees</td>
<td>Persons</td>
</tr>
<tr>
<td>CDs, Service</td>
<td>Sells</td>
</tr>
<tr>
<td>Business plan</td>
<td>Release</td>
</tr>
<tr>
<td>Certified</td>
<td>Oracle, SAP, etc.</td>
</tr>
<tr>
<td>Beginners</td>
<td>Preferred by</td>
</tr>
<tr>
<td>Rpm</td>
<td>Packages</td>
</tr>
<tr>
<td>Market</td>
<td>Customisation</td>
</tr>
</tbody>
</table>
Customising Debian

- Debian > 20000 packages
- Focus on *medical subset* of those packages
- Easy installation and configuration
- Automatic installation → cloud computing
- Maintaining a general infrastructure for medical users
- Propagate the idea of Free Software in medicine
- Completely integrated into Debian - **no fork**

*Basic idea: Do not make a separate distribution but make Debian fit for medical care instead*
Customising Debian

- Debian > 20000 packages
- Focus on *medical subset* of those packages
- Easy installation and configuration
- Automatic installation → cloud computing
- Maintaining a general infrastructure for medical users
- Propagate the idea of Free Software in medicine
- Completely integrated into Debian - no fork

*Basic idea: Do not make a separate distribution but make Debian fit for medical care instead*
Customising Debian

- Debian > 20000 packages
- Focus on *medical subset* of those packages
- Easy installation and configuration
  - Automatic installation → cloud computing
- Maintaining a general infrastructure for medical users
- Propagate the idea of Free Software in medicine
- Completely integrated into Debian - **no fork**

*Basic idea: Do not make a separate distribution but make Debian fit for medical care instead*
Customising Debian

- Debian > 20000 packages
- Focus on *medical subset* of those packages
- Easy installation and configuration
- Automatic installation → cloud computing
- Maintaining a general infrastructure for medical users
- Propagate the idea of Free Software in medicine
- Completely integrated into Debian - **no fork**

**Basic idea:** Do not make a separate distribution but make Debian fit for medical care instead
Customising Debian

- Debian > 20000 packages
- Focus on *medical subset* of those packages
- Easy installation and configuration
- Automatic installation $\rightarrow$ cloud computing
- Maintaining a general infrastructure for medical users
- Propagate the idea of Free Software in medicine
- Completely integrated into Debian - **no fork**

*Basic idea: Do not make a separate distribution but make Debian fit for medical care instead*
Customising Debian

- Debian > 20000 packages
- Focus on *medical subset* of those packages
- Easy installation and configuration
- Automatic installation → cloud computing
- Maintaining a general infrastructure for medical users
- Propagate the idea of Free Software in medicine
- Completely integrated into Debian - no fork

Basic idea: Do not make a separate distribution but make Debian fit for medical care instead
Customising Debian

- Debian > 20000 packages
- Focus on *medical subset* of those packages
- Easy installation and configuration
- Automatic installation → cloud computing
- Maintaining a general infrastructure for medical users
- Propagate the idea of Free Software in medicine
- Completely integrated into Debian - no fork

*Basic idea: Do not make a separate distribution but make Debian fit for medical care instead.*
Customising Debian

- Debian > 20000 packages
- Focus on *medical subset* of those packages
- Easy installation and configuration
- Automatic installation \(\rightarrow\) cloud computing
- Maintaining a general infrastructure for medical users
- Propagate the idea of Free Software in medicine
- Completely integrated into Debian - no fork

**Basic idea:** Do not make a separate distribution but make Debian fit for medical care instead
Debian - adaptable for any purpose?

- Developed by about 1000 volunteers
- Flexible, not bound on commercial interest
- Strict rules (policy) glue all things together
- Common interest of each individual developer: Get the best operating system for himself.
- Developers have children in real life or work in the field of medicine etc.
- In contrast to employees of companies every single Debian developer has the freedom and ability to realize his vision
- Every developer is able to influence the development of Debian - he just has to do it.

*Do-O-Cracy = "The doer decides"*
Debian - adaptable for any purpose?

- Developed by about 1000 volunteers
- Flexible, not bound on commercial interest
- Strict rules (policy) glue all things together
- Common interest of each individual developer: Get the best operating system for himself.
- Developers have children in real life or work in the field of medicine etc.
- In contrast to employees of companies every single Debian developer has the freedom and ability to realize his vision
- Every developer is able to influence the development of Debian - he just has to do it.

Do-O-Cracy = "The doer decides"
Debian - adaptable for any purpose?

- Developed by about 1000 volunteers
- Flexible, not bound on commercial interest
- Strict rules (policy) glue all things together
- Common interest of each individual developer: Get the best operating system for himself.
- Developers have children in real life or work in the field of medicine etc.
- In contrast to employees of companies every single Debian developer has the freedom and ability to realize his vision
- Every developer is able to influence the development of Debian - he just has to do it.

Do-O-Cracy = "The doer decides"
Debian - adaptable for any purpose?

- Developed by about 1000 volunteers
- Flexible, not bound on commercial interest
- Strict rules (policy) glue all things together
- Common interest of each individual developer: Get the best operating system for himself.

Developers have children in real life or work in the field of medicine etc.

In contrast to employees of companies every single Debian developer has the freedom and ability to realize his vision.

Every developer is able to influence the development of Debian - he just has to *do* it.

*Do-O-Cracy = "The doer decides"*
Debian - adaptable for any purpose?

- Developed by about 1000 volunteers
- Flexible, not bound on commercial interest
- Strict rules (policy) glue all things together
- Common interest of each individual developer: Get the best operating system for himself.
- Developers have children in real life or work in the field of medicine etc.
- In contrast to employees of companies every single Debian developer has the freedom and ability to realize his vision
- Every developer is able to influence the development of Debian - he just has to do it.

*Do-O-Cracy = "The doer decides"*
Debian - adaptable for any purpose?

- Developed by about 1000 volunteers
- Flexible, not bound on commercial interest
- Strict rules (policy) glue all things together
- Common interest of each individual developer: Get the best operating system for himself.
- Developers have children in real life or work in the field of medicine etc.
- In contrast to employees of companies every single Debian developer has the freedom and ability to realize his vision
- Every developer is able to influence the development of Debian - he just has to do it.

Do-O-Cracy = "The doer decides"
Debian - adaptable for any purpose?

- Developed by about 1000 volunteers
- Flexible, not bound on commercial interest
- Strict rules (policy) glue all things together
- Common interest of each individual developer: Get the best operating system for himself.
- Developers have children in real life or work in the field of medicine etc.
- In contrast to employees of companies every single Debian developer has the freedom and ability to realize his vision
- Every developer is able to influence the development of Debian - he just has to do it.

Do-O-Cracy = "The doer decides"
Debian - adaptable for any purpose?

- Developed by about 1000 volunteers
- Flexible, not bound on commercial interest
- Strict rules (policy) glue all things together
- Common interest of each individual developer: Get the best operating system for himself.
- Developers have children in real life or work in the field of medicine etc.
- In contrast to employees of companies every single Debian developer has the freedom and ability to realize his vision
- Every developer is able to influence the development of Debian - he just has to do it.

*Do-O-Cracy* = "The doer decides"
1 Debian Med
   - Debian Pure Blend for medical care and health science
   - Why Debian

2 Implementation
   - Available packages
   - Biological databases

3 Looking beyond
   - Alternatives and prospectus
Programming language support

*BioPerl* Collection of Perl tools for computational molecular biology

*BioPython* Python library for computational molecular biology

*BioRuby* Ruby tools for computational molecular biology

*BioJava* Java API to biological data and applications

*BioSQUID* library of C code functions for sequence analysis
Programming language support

**BioPerl** Collection of Perl tools for computational molecular biology

**BioPython** Python library for computational molecular biology

**BioRuby** Ruby tools for computational molecular biology

**BioJava** Java API to biological data and applications

**BioSQUID** Library of C code functions for sequence analysis
Programming language support

**BioPerl** Collection of Perl tools for computational molecular biology

**BioPython** Python library for computational molecular biology

**BioRuby** Ruby tools for computational molecular biology

**BioJava** Java API to biological data and applications

**BioSQUID** library of C code functions for sequence analysis
Programming language support

- **BioPerl**: Collection of Perl tools for computational molecular biology
- **BioPython**: Python library for computational molecular biology
- **BioRuby**: Ruby tools for computational molecular biology
- **BioJava**: Java API to biological data and applications
- **BioSQUID**: Library of C code functions for sequence analysis
Programming language support

<table>
<thead>
<tr>
<th><strong>BioPerl</strong></th>
<th>Collection of Perl tools for computational molecular biology</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BioPython</strong></td>
<td>Python library for computational molecular biology</td>
</tr>
<tr>
<td><strong>BioRuby</strong></td>
<td>Ruby tools for computational molecular biology</td>
</tr>
<tr>
<td><strong>BioJava</strong></td>
<td>Java API to biological data and applications</td>
</tr>
<tr>
<td><strong>BioSQUID</strong></td>
<td>Library of C code functions for sequence analysis</td>
</tr>
</tbody>
</table>
Widely used software

**BLAST2** Basic Local Alignment Search Tool
official NCBI version of this famous sequence alignment program (Note that databases are not included in Debian; they must be retrieved manually.)

**EMBOSS** European Molecular Biology Open Software Suite
EMBOSS is a free Open Source software analysis package specially developed for the needs of the molecular biology (e.g. EMBnet) user community
**Widely used software**

*BLAST2*  
Basic Local Alignment Search Tool  
official NCBI version of this famous sequence alignment program (Note that databases are not included in Debian; they must be retrieved manually.)

*EMBOSS*  
European Molecular Biology Open Software Suite  
EMBOSS is a free Open Source software analysis package specially developed for the needs of the molecular biology (e.g. EMBnet) user community
Statistics using GNU R

**R-cran-genetics** GNU R package for population genetics
The package provides a library for the statistics environment R that contains classes to represent genotypes and haplotypes at single markers up to multiple markers on multiple chromosomes.

**R-cran-haplo.stats** GNU R package for haplotype analysis
The package provides routines for the GNU R statistics environment for statistical Analysis of indirectly measured Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous.

**Bioconductor** GNU R tools for the analysis and comprehension of genomic data.
Not yet packaged for Debian but work in progress to automate packaging of CRAN and Bioconductor packages.
Statistics using GNU R

**R-cran-genetics** GNU R package for population genetics
The package provides a library for the statistics environment R that contains classes to represent genotypes and haplotypes at single markers up to multiple markers on multiple chromosomes.

**R-cran-haplo.stats** GNU R package for haplotype analysis
The package provides routines for the GNU R statistics environment for statistical Analysis of indirectly measured Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous

**Bioconductor** GNU R tools for the analysis and comprehension of genomic data.
Not yet packaged for Debian but work in progress to automate packaging of CRAN and Bioconductor packages.
Statistics using GNU R

(R-cran-genetics) GNU R package for population genetics
The package provides a library for the statistics environment R that contains classes to represent genotypes and haplotypes at single markers up to multiple markers on multiple chromosomes.

(R-cran-haplo.stats) GNU R package for haplotype analysis
The package provides routines for the GNU R statistics environment for statistical Analysis of indirectly measured Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous

(Bioconductor) GNU R tools for the analysis and comprehension of genomic data.
Not yet packaged for Debian but work in progress to automate packaging of CRAN and Bioconductor packages.
Statistics using GNU R

**R-cran-genetics** GNU R package for population genetics
The package provides a library for the statistics environment R that contains classes to represent genotypes and haplotypes at single markers up to multiple markers on multiple chromosomes.

**R-cran-haplo.stats** GNU R package for haplotype analysis
The package provides routines for the GNU R statistics environment for statistical Analysis of indirectly measured Haplotypes with Traits and Covariates when Linkage Phase is Ambiguous

**Bioconductor** GNU R tools for the analysis and comprehension of genomic data.
Not yet packaged for Debian but work in progress to automate packaging of CRAN and Bioconductor packages.

There are some more general R packages recommended by

**med-statistics**

Andreas Tille (Libre Software Meeting) Sequence analysis and bioinformatics LSM, Nantes 2009 16 / 32
Phylogenetic analysis

- **Altree**: Perform phylogeny based analyses
- **fastdnaml**: Construction of phylogenetic trees of DNA sequences
- **Njplot**: Phylogenetic tree drawing program
- **Tree-puzzle**: Reconstruction of phylogenetic trees by maximum likelihood
- **Treeviewx**: Displays and prints phylogenetic trees
Phylogenetic analysis

- **Altree**: Perform phylogeny based analyses
- **fastdnaml**: Construction of phylogenetic trees of DNA sequences
- **Njplot**: Phylogenetic tree drawing program
- **Tree-puzzle**: Reconstruction of phylogenetic trees by maximum likelihood
- **Treeviewx**: Displays and prints phylogenetic trees
- **Phylip**: Package of programs for inferring phylogenies
- **Treetool**: Interactive tool for displaying phylogenetic trees
Genetics and analysis of RNA sequences

Genetics:

**Fastlink**  Faster version of pedigree programs of Linkage

**Loki**  MCMC linkage analysis on general pedigrees

**Plink**  Whole-genome association analysis toolset

**R-cran-qtl**  GNU R package for genetic marker linkage analysis

Analysis of RNA sequences:

**Infernal**  Inference of RNA secondary structural alignments

**Rnahybrid**  Fast and effective prediction of microRNA/target duplexes
Genetics and analysis of RNA sequences

Genetics:
- **Fastlink**: Faster version of pedigree programs of Linkage
- **Loki**: MCMC linkage analysis on general pedigrees
- **Plink**: Whole-genome association analysis toolset
- **R-cran-qtl**: GNU R package for genetic marker linkage analysis

Analysis of RNA sequences:
- **Infernal**: Inference of RNA secondary structural alignments
- **Rnahybrid**: Fast and effective prediction of microRNA/target duplexes
<table>
<thead>
<tr>
<th>Program</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>amap-alig</strong></td>
<td>Protein multiple alignment by sequence annealing</td>
</tr>
<tr>
<td><strong>Boxshade</strong></td>
<td>Pretty-printing of multiple sequence alignments</td>
</tr>
<tr>
<td><strong>Dialign(-tx)</strong></td>
<td>Segment-based multiple sequence alignment</td>
</tr>
<tr>
<td><strong>Exonerate</strong></td>
<td>Generic tool for pairwise sequence comparison</td>
</tr>
<tr>
<td><strong>Gff2aplot</strong></td>
<td>Pair-wise alignment-plots for genomic sequences in PostScript</td>
</tr>
<tr>
<td><strong>Hmmer</strong></td>
<td>Profile hidden Markov models for protein sequence analysis</td>
</tr>
<tr>
<td><strong>Kalign</strong></td>
<td>Global and progressive multiple sequence alignment</td>
</tr>
<tr>
<td><strong>Mafft</strong></td>
<td>Multiple alignment program for amino acid or nucleotide sequences</td>
</tr>
<tr>
<td><strong>Mummer</strong></td>
<td>Efficient sequence alignment of full genomes</td>
</tr>
<tr>
<td><strong>Muscle</strong></td>
<td>Multiple alignment program of protein sequences</td>
</tr>
</tbody>
</table>
Sequence alignments and related programs (cont.)

- **Poa**: Partial Order Alignment for multiple sequence alignment
- **Probcons**: PROBabilistic CONSistency-based multiple sequence alignment
- **Proda**: Multiple alignment of protein sequences
- **Seaview**: Multiplatform interface for sequence alignment and phylogeny
- **Sibsim4**: Align expressed RNA sequences on a DNA template
- **Sigma-align**: Simple greedy multiple alignment of non-coding DNA sequences
- **Sim4**: Tool for aligning cDNA and genomic DNA
- **T-coffee**: Multiple Sequence Alignment
- **Wise**: Comparison of biopolymers, commonly DNA and protein sequences
Molecular modelling and molecular dynamics

Adun.app  Molecular Simulator for GNUstep
Autogrid  Pre-calculate binding of ligands to their receptor
Gamgi    Construct, view and analyse atomic structures
Garlic   Visualisation program for biomolecules
Gdpc     Visualiser of molecular dynamic simulations
Ghemical GNOME molecular modelling environment
Gromacs  Molecular dynamics simulator, with building and analysis tools
Pymol    Molecular Graphics System
R-other-bio3d  GNU R package for biological structure analysis
Rasmol   Visualise biological macromolecules
### Molecular modelling and molecular dynamics

<table>
<thead>
<tr>
<th><strong>Software</strong></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Adun.app</strong></td>
<td>Molecular Simulator for GNUstep</td>
</tr>
<tr>
<td><strong>Autogrid</strong></td>
<td>Pre-calculate binding of ligands to their receptor</td>
</tr>
<tr>
<td><strong>Gamgi</strong></td>
<td>Construct, view and analyse atomic structures</td>
</tr>
<tr>
<td><strong>Garlic</strong></td>
<td>Visualisation program for biomolecules</td>
</tr>
<tr>
<td><strong>Gdpc</strong></td>
<td>Visualiser of molecular dynamic simulations</td>
</tr>
<tr>
<td><strong>Ghemical</strong></td>
<td>GNOME molecular modelling environment</td>
</tr>
<tr>
<td><strong>Gromacs</strong></td>
<td>Molecular dynamics simulator, with building and analysis tools</td>
</tr>
<tr>
<td><strong>Pymol</strong></td>
<td>Molecular Graphics System</td>
</tr>
<tr>
<td><strong>R-other-bio3d</strong></td>
<td>GNU R package for biological structure analysis</td>
</tr>
<tr>
<td><strong>Rasmol</strong></td>
<td>Visualise biological macromolecules</td>
</tr>
<tr>
<td><strong>Autodocktools</strong></td>
<td>GUI to help set up, launch and analyse AutoDock dockings</td>
</tr>
</tbody>
</table>
High-throughput sequencing

- “Next-generation sequencing”
- Chip-systems to sequence a genome
- Reads are very short (40 nucleotides rather than traditionally about 600)
- Enormous amount of chromosomal regions covered

Last-align: Genome-scale comparison of biological sequences
Maq: Maps short fixed-length polymorphic DNA sequence reads to reference sequences
Ssake: Genomics application for assembling millions of very short DNA sequences
Velvet: Nucleic acid sequence assembler for very short reads
Mikrobiological packages

- More than 80 Packages
- Overview at according tasks page of Debian Med project
- Software developed by
  - National Center for Biotechnology Information (NCBI)
  - Sanger Institute
  - The Institute for Genomic Research (TIGR)
  - . . .
Mikrobiological packages

- More than 80 Packages
- Overview at *according tasks page of Debian Med project*

Software developed by

- National Center for Biotechnology Information (NCBI)
- Sanger Institute
- The Institute for Genomic Research (TIGR)
- . . .
Mikrobiological packages

- More than 80 Packages
- Overview at
  according tasks page of Debian Med project
- Software developed by
  - National Center for Biotechnology Information (NCBI)
  - Sanger Institute
  - The Institute for Genomic Research (TIGR)
  - ...
Mikrobiological packages

- More than 80 Packages
- Overview at according tasks page of Debian Med project
- Software developed by
  - National Center for Biotechnology Information (NCBI)
  - Sanger Institute
  - The Institute for Genomic Research (TIGR)
  - ...
Mikrobiological packages

- More than 80 Packages
- Overview at according tasks page of Debian Med project
- Software developed by
  - National Center for Biotechnology Information (NCBI)
  - Sanger Institute
  - The Institute for Genomic Research (TIGR)
  - ...
Mikrobiological packages

- More than 80 Packages
- Overview at [according tasks page of Debian Med project](#)
- Software developed by
  - National Center for Biotechnology Information (NCBI)
  - Sanger Institute
  - The Institute for Genomic Research (TIGR)
  - ...
Mikrobiological packages

- More than 80 Packages
- Overview at according tasks page of Debian Med project
- Software developed by
  - National Center for Biotechnology Information (NCBI)
  - Sanger Institute
  - The Institute for Genomic Research (TIGR)
  - ...

Andreas Tille (Libre Software Meeting)
Sequence analysis and bioinformatics
LSM, Nantes 2009
23 / 32
### DebTags

```sql
udd=# SELECT tag, COUNT(*) FROM debtags
    WHERE tag LIKE '%bio%'
    GROUP BY tag ORDER BY tag;
```

<table>
<thead>
<tr>
<th>tag</th>
<th>count</th>
</tr>
</thead>
<tbody>
<tr>
<td>biology::emboss</td>
<td>2</td>
</tr>
<tr>
<td>biology::format:aln</td>
<td>9</td>
</tr>
<tr>
<td>biology::format:fasta</td>
<td>9</td>
</tr>
<tr>
<td>biology::nuceleic-acids</td>
<td>11</td>
</tr>
<tr>
<td>biology::peptidic</td>
<td>12</td>
</tr>
<tr>
<td>field::biology</td>
<td>174</td>
</tr>
<tr>
<td>field::biology:bioinformatics</td>
<td>86</td>
</tr>
<tr>
<td>field::biology:molecular</td>
<td>8</td>
</tr>
<tr>
<td>field::biology:structural</td>
<td>16</td>
</tr>
</tbody>
</table>

(9 rows)
1. Debian Med
   - Debian Pure Blend for medical care and health science
   - Why Debian

2. Implementation
   - Available packages
   - Biological databases

3. Looking beyond
   - Alternatives and prospectus
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate
- Solution also works for astronomy and meteorology

**getData**

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
    - Moving target: Stable distribution will be out of date soon
    - Remote service seems appropriate
- Solution also works for astronomy and meteorology

**getData**

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate
- Solution also works for astronomy and meteorology

getData

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwiths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate

- Solution also works for astronomy and meteorology

**getData**

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate

- Solution also works for astronomy and meteorology

**getData**
- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate
- Solution also works for astronomy and meteorology

**getData**

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate
- Solution also works for astronomy and meteorology

**getData**

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate

- Solution also works for astronomy and meteorology

---

dataGet

- Obtain data from external source
- Move data to local mirror
  - Preparation of configuration file for particular system that deals with the database
  - getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate
- Solution also works for astronomy and meteorology

**getData**

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
  - getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate

- Solution also works for astronomy and meteorology

**getData**

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
  - People are much welcome to join this development (Google Summer of Code project)
How to install large databases

- Bundling into Debian package makes no sense
  - Size costs bandwidths and mirror space
  - Moving target: Stable distribution will be out of date soon
  - Remote service seems appropriate
- Solution also works for astronomy and meteorology

**getData**

- Obtain data from external source
- Move data to local mirror
- Preparation of configuration file for particular system that deals with the database
- getData is still in a proof of concept stage
- People are much welcome to join this development (Google Summer of Code project)
Open Database License (ODbL) v1.0

- **Open Data Commons**
- License agreement intended to allow users to freely share, modify, and use this Database while maintaining this same freedom for others
- Databases can contain a wide variety of types of content (images, audiovisual material, and sounds all in the same database, for example)
- ODbL only governs the rights over the Database, and not the contents of the Database individually
Open Database License (ODbL) v1.0

- **Open Data Commons**
- License agreement intended to allow users to freely share, modify, and use this Database while maintaining this same freedom for others
- Databases can contain a wide variety of types of content (images, audiovisual material, and sounds all in the same database, for example)
- ODbL only governs the rights over the Database, and not the contents of the Database individually
**Open Data Commons**

License agreement intended to allow users to freely share, modify, and use this Database while maintaining this same freedom for others.

Databases can contain a wide variety of types of content (images, audiovisual material, and sounds all in the same database, for example).

ODbL only governs the rights over the Database, and not the contents of the Database individually.
Open Database License (ODbL) v1.0

Open Data Commons

License agreement intended to allow users to freely share, modify, and use this Database while maintaining this same freedom for others.

Databases can contain a wide variety of types of content (images, audiovisual material, and sounds all in the same database, for example).

ODbL only governs the rights over the Database, and not the contents of the Database individually.
Debian Med
- Debian Pure Blend for medical care and health science
- Why Debian

Implementation
- Available packages
- Biological databases

Looking beyond
- Alternatives and prospectus
Alternatives

BioLinux

- Based on Debian
- Create a policy incompatible structure in `/usr/local/biolinux`
- Some software not yet available in Debian but really sloppy with licenses
- We try to include the missing stuff in Debian to create a policy compliant, really free system
- Hope BioLinux people will adopt this

FreeBSD ports collection Biology

- Also contains a fair amount of biological software
- Only a few unimportant missing in Debian
Alternatives

**BioLinux**

- Based on Debian
- Create a policy incompatible structure in `/usr/local/biolinux`
- Some software not yet available in Debian but really sloppy with licenses
- We try to include the missing stuff in Debian to create a policy compliant, really free system
- Hope BioLinux people will adopt this

**FreeBSD ports collection Biology**

- Also contains a fair amount of biological software
- Only a few unimportant missing in Debian
## Alternatives

### BioLinux
- Based on Debian
- Create a policy incompatible structure in `/usr/local/biolinux`
- Some software not yet available in Debian but really sloppy with licenses
- We try to include the missing stuff in Debian to create a policy compliant, really free system
- Hope BioLinux people will adopt this

### FreeBSD ports collection Biology
- Also contains a fair amount of biological software
- Only a few unimportant missing in Debian
There are good requisites in Debian

- Most important tools of molecular biology, structural biology and bioinformatics for use in life sciences are included
- Further increase of interest of developers and users and getting them involved in the project
- Turning Debian into the distribution of choice for people working in the field of medicine because there is best support for free medical software
Prospectus

- There are good requisites in Debian
- Most important tools of molecular biology, structural biology and bioinformatics for use in life sciences are included
- Further increase of interest of developers and users and getting them involved in the project
- Turning Debian into the distribution of choice for people working in the field of medicine because there is best support for free medical software
There are good requisites in Debian

- Most important tools of molecular biology, structural biology and bioinformatics for use in life sciences are included
- Further increase of interest of developers and users and getting them involved in the project
- Turning Debian into the distribution of choice for people working in the field of medicine because there is best support for free medical software
Prospectus

- There are good requisites in Debian
- Most important tools of molecular biology, structural biology and bioinformatics for use in life sciences are included
- Further increase of interest of developers and users and getting them involved in the project
- Turning Debian into the distribution of choice for people working in the field of medicine because there is best support for free medical software
This talk is available at
http://people.debian.org/~tille/talks/
Andreas Tille <tille@debian.org>