

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

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

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?

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?

Medical imaging

- Debian Med can only include *existing* software
- Fair amount of high quality Free Software for medical imaging
 - Aeskulap, Amide: Medical image viewers
 - Dcmtnk: 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](#)

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](#)

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](#)

Medical imaging

- Debian Med can only include *existing* software
- Fair amount of high quality Free Software for medical imaging
 - Aeskulap, Amide: Medical image viewers
 - Dcmtnk: 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
 - Dcmtnk: 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

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

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

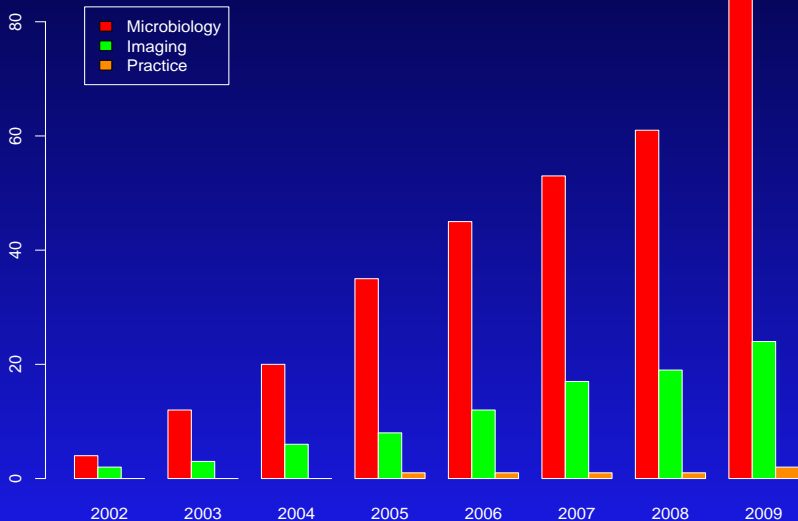
- Most established branch of Debian Med because of good coverage by upstream software
- Fostering
 - Development at universities
 - Organised funding
- Hindering
 - Advertising for proprietary software
 - Different preferences of initiators

Molecular and structural biology, bioinformatics

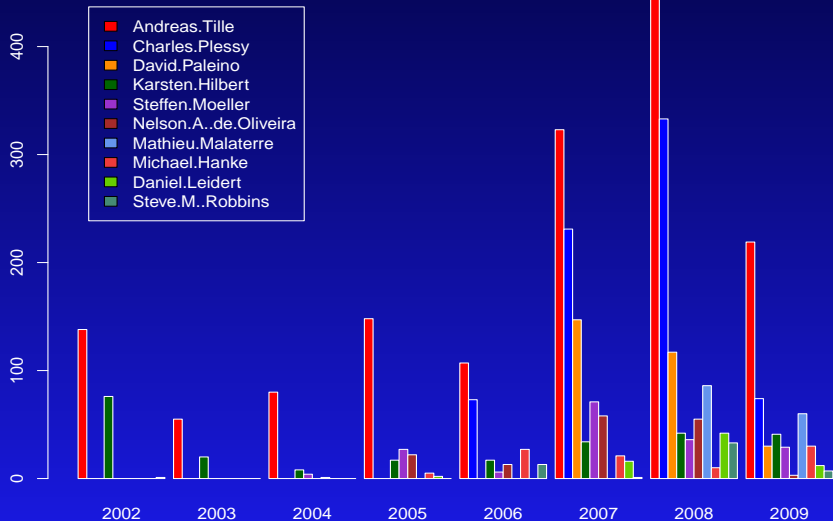
- Most established branch of Debian Med because of good coverage by upstream software
- Fostering
 - Development at universities
 - Organised funding
- Hindering
 - Advertising for proprietary software
 - Different preferences of initiators

Selected metapackages of Debian Med

Number of dependencies of selected Debian Med metapackages



Top 10 posters on debian-med@lists.debian.org



- 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

Debian

Company

Structure

Organisation

Differences to commercial distributions

Commercial distributor

Debian

Company

Structure

Organisation

Employees

Persons

Volunteers

Differences to commercial distributions

Commercial distributor

Debian

Company

Structure

Organisation

Employees

Persons

Volunteers

CDs, Service

Sells

Nothing

Differences to commercial distributions

Commercial distributor

Debian

Company

Structure

Organisation

Employees

Persons

Volunteers

CDs, Service

Sells

Nothing

Business plan

Release

If 0 RC-bugs

Differences to commercial distributions

Commercial distributor

Debian

Company

Structure

Organisation

Employees

Persons

Volunteers

CDs, Service

Sells

Nothing

Business plan

Release

If 0 RC-bugs

Certified

Oracle, SAP, etc.

Runs in principle

Differences to commercial distributions

Commercial distributor

Debian

Company

Structure

Organisation

Employees

Persons

Volunteers

CDs, Service

Sells

Nothing

Business plan

Release

If 0 RC-bugs

Certified

Oracle, SAP, etc.

Runs in principle

Beginners

Preferred by

Administrators

Differences to commercial distributions

Commercial distributor

Debian

Company

Structure

Organisation

Employees

Persons

Volunteers

CDs, Service

Sells

Nothing

Business plan

Release

If 0 RC-bugs

Certified

Oracle, SAP, etc.

Runs in principle

Beginners

Preferred by

Administrators

Rpm

Packages

Deb

Differences to commercial distributions

Commercial distributor		Debian
Company	Structure	Organisation
Employees	Persons	Volunteers
CDs, Service	Sells	Nothing
Business plan	Release	If 0 RC-bugs
Certified	Oracle, SAP, etc.	Runs in principle
Beginners	Preferred by	Administrators
Rpm	Packages	Deb
Market	Customisation	Do-O-Cracy

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 → 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*

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

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

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

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-qt1 GNU R package for genetic marker linkage analysis

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-qt1 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

Sequence alignments and related programs

amap-align Protein multiple alignment by sequence annealing

Boxshade Pretty-printing of multiple sequence alignments

Dialign(-tx) Segment-based multiple sequence alignment

Exonerate Generic tool for pairwise sequence comparison

Gff2aplot Pair-wise alignment-plots for genomic sequences in PostScript

Hmmer Profile hidden Markov models for protein sequence analysis

Kalign Global and progressive multiple sequence alignment

Mafft Multiple alignment program for amino acid or nucleotide sequences

Mummer Efficient sequence alignment of full genomes

Muscle Multiple alignment program of protein sequences

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

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

Ghchemical 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

Autodocktools GUI to help set up, launch and analyse AutoDock dockings

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 \(TIGR\)*](#)
 - ...

DebTags

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

tag	count
biology::emboss	2
biology::format:aln	9
biology::format:fasta	9
biology::nuceleic-acids	11
biology::peptidic	12
field::biology	174
field::biology:bioinformatics	86
field::biology:molecular	8
field::biology:structural	16

(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 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

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

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 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

- 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

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

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

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

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

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

A landscape photograph of a dry, eroded valley. In the foreground, a large, smooth, brown rock sits on a sandy slope. The middle ground features a small, irregular pond of turquoise water, surrounded by eroded, light-brown earth. The background shows rolling hills under a dramatic sky with large, white and grey clouds. A faint rainbow is visible in the upper right portion of the sky, with light rays extending downwards.

This talk is available at
<http://people.debian.org/~tille/talks/>
Andreas Tille <tille@debian.org>