

Bioinformatik und medizinische Bildverarbeitung

Mit Debian Med Aufgaben in Bioinformatik und der
medizinischen Bildverarbeitung lösen

Andreas Tille

Debian

Chemnitz, 12. März 2017

1 / 38

- 1 Historisches
- 2 Debian Pure Blends
- 3 Testen für wissenschaftliche Anwendung
- 4 Support
- 5 Bioinformatik
- 6 Medizinische Bildverarbeitung
- 7 Ausblick

2 / 38

Bordeaux 2001 DebConf 1



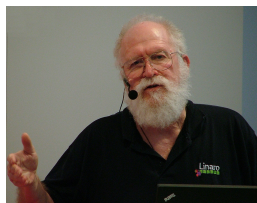
3 / 38

Vorbereitung erster Vortrag über Debian Med



4 / 38

Lektion von DebConf 15 Heidelberg



I After 15 years people are starting to use what I'm
working on.

John Only 15 years? You young guys should be more
patient. I should hurry up since after 15 years I
might be dead but you have so much time.

Image by Olaf Kosinsky CC BY-SA 3.0

5 / 38

Motivation

Weil es möglich ist → Do-ocracy

6 / 38

Was ist ein "Paket"?

Ein Paket ist ein Archiv von Dateien und Metadaten:

- sofort lauffähige Software
- Metainformation zur Definition von Abhängigkeiten
- Dokumentation
- Beispiele für die Anwendung

→ Installation per Mausklick auf beliebiger Anzahl von Computern

Spezifische Lösungen innerhalb Debian

- Debian > 50000 Pakete
- Biologen & Mediziner brauchen nur sehr kleine spezifische *Untermenge*
- Leichte Installation + Konfiguration dieser Untermenge
- Automatische Installation → Cloud
- Dokumentation und Metadaten zur Software (Publikationen, Screenshots, Übersetzungen)
- Vollständig integriert in Debian → **kein „Fork“**

*Grundidee: Keine separate Distribution
sondern Debian für spezifische Lösungen in
Medizin und Biologie anpassen*

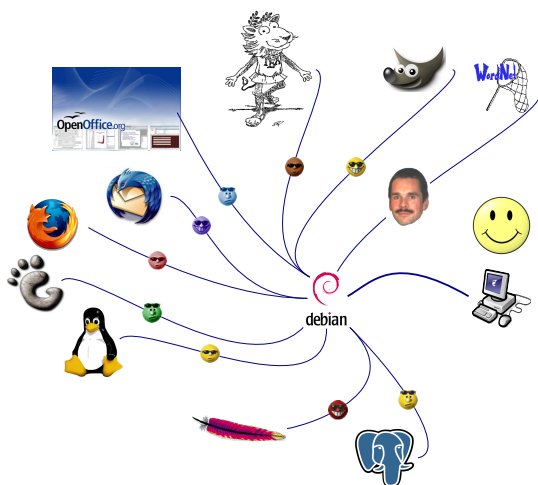
Abhängigkeiten

- Software hängt in der Regel von der Verfügbarkeit anderer Software ab
- Beispiele:
 - Java-Run-Time, gemeinsam genutzter Code (sogenannte Bibliotheken)
 - Bioinformatik-Pipeline hängt ab von Assembler, Mapper, etc.
- Paketiersystem sichert die Verfügbarkeit aller Abhängigkeiten ohne zusätzliche Maßnahmen
- ... zumindest bei modernen Linux Distributionen

Sofort einsetzbare Software

- erzeugt aus dem Quellcode
- ermöglicht wissenschaftliche Validierung
- getestet zum Zeitpunkt der Paketerstellung
- regelmäßige Tests zur Reproduktion erwarteter Resultate
- ... zumindest bei der Verwendung von Debian Paketen

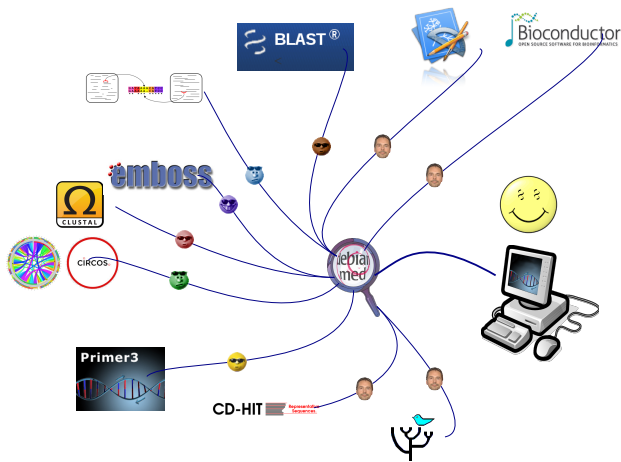
Debian



Bioinformatik in Debian

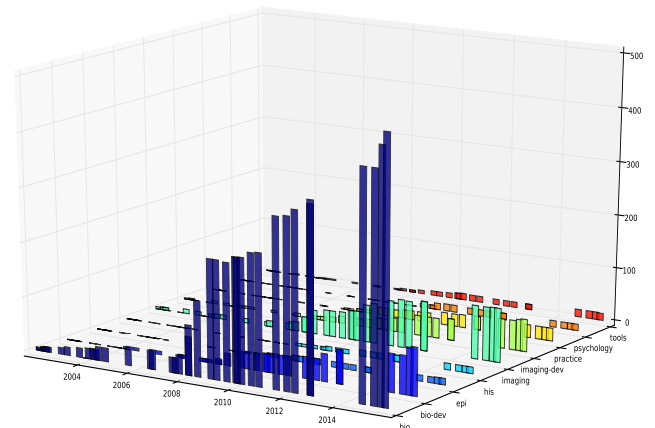


Med-bio task of Debian Med



13 / 38

Pakete in ausgewählten Tasks von Debian Med



14 / 38

Getestete Software in der Wissenschaft

- Freie Software ermöglicht die Untersuchung des Algorithmus
- Software wird zum Zeitpunkt der Erstellung des Binärcodes getestet
- Pakete werden automatisch installiert und regelmäßig getestet
- Identischer Code wird weltweit von Wissenschaftlern benutzt, die Probleme berichten können
- Paketsystem erstellt Protokoll zu welcher Zeit welche Softwareversion installiert war

15 / 38

Eigene Tests für reproduzierbare Ergebnisse

- Erstellen eigener Testsuites möglich
- Vergleich durch prozessieren von Standarddaten mit erwarteten Ergebnissen
- Test in virtueller Maschine (oder chroot) vor dem Ausrollen

16 / 38

Debian Med organisiert GSoC Projekt

- Google Summer of Code / Outreachy
- Bioinformatikstudenten schreiben Tests
- Bevorzugung vielgenutzter Debian Pakete (Popularity Contest)
- Während des Testens Fehler erkannt und behoben

17 / 38

Automatische Versionsüberwachung

- Verschiedene eingesetzte Tools sind veraltet
- Bewußt oder Unwissen?
- Per *uscan* automatisch informieren lassen

18 / 38

Support in Freien Software Teams

*Am Morgen aufwachen und feststellen,
daß jemand anderes Dein Problem
von gestern Abend gelöst hat.*

19 / 38

Illusion? – Ein Beispiel (giira)

Sat, 8 Feb 2014 00:16 A. Tille
Need some help for Java package

Sat, 8 Feb 2014 08:15 O. Sallou
I gonna have a look

Sat, 8 Feb 2014 08:42 O. Sallou
I was close to get it

Sat, 8 Feb 2014 09:11 O. Sallou
Got this problem, detected new one, need to contact authors

Sat, 8 Feb 2014 09:30 A. Tille
Thanks for your effort, I can give the authors a phone call

Sat, 8 Feb 2014 10:57 A. Tille
There is some other minor problem

Sat, 8 Feb 2014 12:44 O. Sallou
There is a simple workaround

Mon, 10 Feb 2014 11:03 O. Sallou
I asked authors and they solved remaining issue

Mon, 10 Feb 2014 13:51 A. Tille
Thanks. Some last advise needed

Mon, 10 Feb 2014 14:02 O. Sallou
Code updated, problem solved

20 / 38

Status Bioinformatik

- Am weitesten entwickelter Zweig von Debian Med
- Sehr große Zahl verfügbarer Freier Software
- Begünstigend
 - Entwicklung an Universitäten und Forschungseinrichtungen
 - Tendenz Freie Software zu nutzen und zu Entwickeln
 - Veröffentlichung von Artikeln + Software
- Nachteilig
 - Industrielle Softwareentwicklung
 - Verständnis Freier Software teilweise nicht ausgeprägt

21 / 38

Viel genutzte Bioinformatikanwendungen

[EMBOSS](#) European Molecular Biology Open Software Suite

[PyMOL](#) Grafiksystem für Moleküle, ausgerichtet auf mittelgroße bis große Biomoleküle wie Proteine

[RasMol](#) Visualisierung biologischer Makromoleküle

[clustalx](#) Multiples Alignment von Nukleotid- und Proteinsequenzen

[gbrowse](#) Generic Genome Browser: konfigurierbarer webbasierter Genombrowser

[NCBI-tools](#) NCBI-Bibliotheken für Biologieprogramme

[Bioconductor limma](#) Lineare Modelle für Microarray Daten

[NJplot](#) Darstellung phylogenetischer Bäume

→ [Bioinformatikanwendungen in Debian](#) (ca. 500)

22 / 38

Phylogenetische Analysen

[NJplot](#) Darstellung phylogenetischer Bäume

[Gubbins](#) Genealogies Unbiased By recomBINations In Nucleotide Sequences

[PhyIip](#) Programmsammlung für phylogenetische Analysen¹

[PhyML](#) Phylogenetische Schätzung per Maximum Likelihood Methode

[seaview](#) Erstellen und Bearbeiten von Alignments und Erstellen von Bäumen

[Treeviewx](#) Darstellung phylogenetischer Bäume

[r-cran-ape](#) GNU R Paket zur Analyse von Phylogenie und Evolution

[MrBayes](#) Bayesian Inference und Modelauswahl zwischen phylogenetischen und evolutionären Modellen

[RaXML](#) Randomized Accelerated Maximum Likelihood phylogenetischer Bäume

[fsa](#) Multiple Alignments durch paarweise Schätzungen der Homologie

[fasttree](#) Phylogenetische Bäume aus Alignments von Nucleotid- oder Proteinsequenzen

→ [Übersicht über mehr Phylogenieprogramme](#)

¹ Bemühungen zur Befreiung von Software

23 / 38

Sequenzierung der dritten Generation (NGS)

[bedtools](#) Sehr umfangreicher Werkzeugkasten für Genomanalyseaufgaben

[BWA](#) Burrows-Wheeler Aligner: Abbilden gering abweichender Sequenzen auf eine große Referenzsequenz

[Bowtie2](#) Ultraschnelles und speichersparendes Alignmentprogramm für kurze DNA-Sequenzen

[LAST](#) Genome-scale comparison of biological sequences

[tabix](#) Generischer Indexer für .tbi und .csi formatierte Dateien

[VCFtools](#) Werkzeugsammlung für die Arbeit mit VCF Dateien

[Maq](#) Mapt kurze polymorphe DNA-Sequenzen mit fester Länge auf Referenzsequenzen

[Velvet](#) DNS Assembler für sehr kurze Reads von Sequenzen

[FASTX-Toolkit](#) Sammlung an Befehlszeilenwerkzeugen zur Vorverarbeitung von kurzen Nukleotidsequenzen

[Ray](#) Paralleler Genome Assembler für parallele DNA-Sequenzierung

→ [Übersicht über mehr Programme für NGS](#)

24 / 38

Sequenzalignment

- [samtools](#) Bearbeiten von Sequenzalignments im SAM und BAM Format
- [ncbi-blast+](#) Basic Local Alignment Search Tool (BLAST) Sequence Werkzeuge vom NCBI
- [clustalw](#) Multiple Alignments von Nukleotid- oder Peptidsequenzen
- [Muscle](#) Programm zur Ausrichtung mehrerer Proteinsequenzen
- [Hmmer](#) Profil-Hidden-Markov-Modelle für Protein-Sequenzanalyse
- [Dialign\(-tx\)](#) Segmentbasiertes multiples Sequenzalignment
- [Exonerate](#) Generisches Werkzeug zum paarweisen Sequenzvergleich
- [Mafft](#) Multiples Alignment Programm für DNA oder Proteinsequenzen
- [amap-align](#) Multiple Alignments durch Sequenz-»annealing« bei Proteinen
- [T-coffee](#) Multiple Sequence Alignment
- [Clustal Omega](#) Universalprogramm für multiple Sequenzalignments (MSA) von Proteinen
- [sim4](#) Werkzeug für Alignments von cDNA mit genomischer DNA

25 / 38

Sequenzalignment (Fortsetzung)

- [Kalign](#) Globales und progressives multiples Sequenzalignment
- [Probcons](#) PROBABilistic CONSistency-based multiple Sequenzalignment
- [Poa](#) Partial Order Alignment für multiple Sequencealignments
- [Mustang](#) MULTiple (protein) STructural AligNment alGorithm
- [Proda](#) Multiple Alignments von Proteinsequenzen
- [GenomicRanges](#) Bioconductor Repräsentation und Manipulation von kurzen Genomalignments
- [Infernal](#) Rückschluss auf Alignments der RNA-Sekundärstruktur
- [Mummer](#) Effizientes Sequenzalignment vollständiger Genome
- [TM-align](#) Proteinstruktur Alignment
- [TopHat](#) Schneller Mapper für Splice Junctions von RNA-Sequenzstücken
- [prank](#) PROBABilistic AligNment Kit für DNA, Codon und DNA Sequenzen
- [Boxshade](#) Pretty-printing von multiplen Sequenzalignments
- [Gff2aplot](#) Paarweise Alignment Plots für Genomsequenzen in PostScript
- [e-PCR](#) Testet DNA Sequenzen auf Vorhandensein von Markern

26 / 38

Exkurs: Wie habe ich die Liste erstellt?

Ultimate Debian Database (UDD)

```
SELECT DISTINCT p.package, p.description, po.vote
FROM packages p
JOIN descriptions d ON
    p.description_md5 = d.description_md5
JOIN popcon po ON p.package = po.package
WHERE maintainer LIKE '%debian-med%'
    AND language = 'en'
    AND p.release = 'sid'
    AND architecture = 'amd64'
    AND (d.description ILIKE '%align%'
        OR long_description ILIKE '%align%')
ORDER BY vote DESC, package;
```

Derzeit liefert diese Abfrage 189 Ergebnisse.

27 / 38

Molekulare Modellierung und Moleküldynamik

- [PyMOL](#) Grafiksystem für Moleküle, ausgerichtet auf mittelgroße bis große Biomoleküle wie Proteine
- [RasMol](#) Visualisierung biologischer Makromoleküle
- [Gromacs](#) Simulator für Moleküldynamik, Werkzeuge für das Erstellen der Modelle und deren Analyse
- [Garlic](#) Darstellungsprogramm für Biomoleküle
- [BALLview](#) Werkzeug für Molekular design und -grafiken
- [Gamgi](#) Konstruktion, Betrachtung und Analyse von Atomstrukturen
- [Gdpc](#) Visualisiert Molekulardynamik-Simulationen
- [Adun.app](#) Molekularsimulator für GNUstep
- [Autogrid](#) Vorberechnung der Bindung von Liganden an ihren Rezeptor
- [Ghemical](#) GNOME Programm zur molekularen Modellierung
- [R-other-bio3d](#) GNU R Paket für Biologische Strukturanalyse

28 / 38

Unterstützung für Entwickler

- [BioPython](#) Python-Werkzeuge für rechnergestützte Molekularbiologie
- [BioPerl](#) Perl-Werkzeuge für etablierte bioinformatische Verfahren
- [BioJava](#) Java API für biologische Daten und Anwendungen
- [BioRuby](#) Ruby Werkzeuge für Molekularbiologie
- [Bioconductor](#) Umfangreiche Paketsammlung für R (mehr Pakete auf Anfrage)
- [Ruffus](#) In Bioinformatik weit verbreitete Computation Pipeline Bibliothek
- [SBML5](#) System Biology Markup Language Bibliothek

→ [Vollständige Übersicht für Entwickler](#)

29 / 38

Alternativen

BioLinux

- Ubuntu Derivat mit extra Bioinformatikpaketen und Desktopanpassungen
- Sehr enge Zusammenarbeit (Entwickler auf Debian Med Sprints)
- Projekt nicht weiterentwickelt - Entwickler empfiehlt Debian Med

OSDDLlinux

- Debian/Ubuntu (?) Derivat zur Medikamentenentwicklung
- Viele Überschneidungen mit Debian Med
- Verschiedene Kontaktversuche zur Zusammenarbeit
- Aktueller Status unklar

FreeBSD ports collection Biology

- Große Anzahl Bioinformatikprogramme
- Einige wenige davon noch nicht für Debian paketiert

30 / 38

Medizinische Bildverarbeitung

- Nutzung in ärztlicher Praxis, Krankenhäusern und wissenschaftlicher Forschung
- NeuroDebian Team betreut großen Teil der Pakete
- Anzahl verfügbarer Freier Software ist geringer als in Bioinformatik
- Abdeckung aller wesentlichen Aufgaben ist gegeben
- Gesucht: Mehr Debian Maintainer für Software auf diesem Gebiet - gern auch Upstream Entwickler
- Mentoring of the Month (MoM)

31 / 38

DICOM Unterstützung

- (X)MedCon Werkzeug zur Konvertierung medizinischer Bilder (DICOM, ECAT, ...)
- GinkgoCADx Kompletter DICOM Betrachter für die ärztliche Praxis
- DCMTK Befehlszeilen-Dienstprogramme der OFFIS-DICOM-Werkzeugsammlung
- orthanc Leichtgewichtiger RESTful DICOM Server für medizinische Bildverarbeitung
- Aeskulap Medizinischer Bildbetrachter und DICOM-Netzwerkclient
- Dicom3tools DICOM-Werkzeuge für Manipulation und Umwandlung medizinischer Bilddateien
- AMIDE Amide's a Medical Imaging Data Examiner
- MRIConvert Konvertierungswerkzeug für medizinische Bilder (DICOM zu NIfTI u.a.)
- gdcm-tools Graswurzel DICOM Bibliothek und Werkzeuge
- ctn Central Test Node, eine DICOM-Implementierung für medizinische Bildverarbeitung
- InVesalius Implementierung von MIDA und konturbasierter Visualisierung

32 / 38

Beispiel: GinkgoCADx

GinkgoCADx enthält einen kompletten DICOM-Betrachter mit fortgeschrittenen Möglichkeiten und Unterstützung für Erweiterungen.

- einfache und anpassbare Oberfläche durch Profile
- vollfunktionale Visualisierung von DICOM-Bildern
- kompletter Werkzeugsatz (messen, markieren, Text, ...)
- unterstützt mehrere Modalitäten (neurologisch, radiologisch, dermatologisch, Augenheilkunde (Ophthalmologie), Ultraschall, Endoskopie, ...)
- DICOM-Konvertierung von JPEG, PNG, GIF und TIFF
- vollständige EMH-Integration: HL7-Standard und IHE-konforme Arbeitsabläufe
- PACS Workstation (C-FIND, C-MOVE, C-STORE...)
- Erweiterbar durch spezielle Anpassungen

33 / 38

Anwendungen medizinischer Bildverarbeitung

- Bildbetrachter für den Arzt
ginkgocadx | aeskulap | amide | dicomscope
- Bioimaging, Volumenberechnungen
imagevis3d, itknap, volview
- Neurologie
ants, fslview, mrtrix, voxbo
- Tomographie
ctsim, mia, odin, plastimatch
- Scanning Probe Microscopy
gwyddion
- Image analysis / Simulation
sofa-apps
- Image format conversion tools (DICOM, MINC, NIfTI)
biosig-tools, dicom3tools, medcon, mriconvert

34 / 38

Entwicklerwerkzeuge

- Allgemein
cimg-dev, gmic, libcv-dev, libvtk6-dev
- Biomedizinische Datenbehandlung
libbiosig-dev, libgdf-dev
- Bildformat Erzeugung und Konvertierung
libmdc2-dev, libminc-dev, libnifti-dev
- Entwicklung spezialisierter Anwendungen:
 - Registrierung und Segmentierung:
libinsighttoolkit4-dev
 - EEG und MEG Anwendungen: *libopenmeeg-dev*
 - 2D und 3D Graustufen Bildbearbeitung: *libmia-dev*
- Programmiersprachen: Python, C, C++, ...

35 / 38

Ausblick

- Debian Med Team hat umfangreiche Sammlung medizinischer Software für das offizielle Debian paketiert
- Ständiger Prozess der Aktualisierung und Neuaufnahme
- Nehmen gerne Paketierwünsche entgegen
- Umfangreiche Metadaten (UDD)
- Erstellung von Tests
- „Befreiung“ von Software
- Wir möchten Debian so attraktiv für die Anwendung in Biologie und Medizin machen, daß informierte Nutzer keine andere Wahl haben, als sich für Debian zu entscheiden.

36 / 38

Folien: → <https://people.debian.org/tille/talks/>
Kontakt: Andreas Tille <tille@debian.org>

Biologische Anwendungen in Debian
Medizinische Bildverarbeitung in Debian