

# Bioinformatik-Webdienste

und ihre tatsächliche Nutzung durch Forscher in  
den Lebenswissenschaften

Universität Tübingen – TL-Stiftung      8. Juli 2008

Sebastian J. Schultheiß      [sebi@tuebingen.mpg.de](mailto:sebi@tuebingen.mpg.de)



EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN

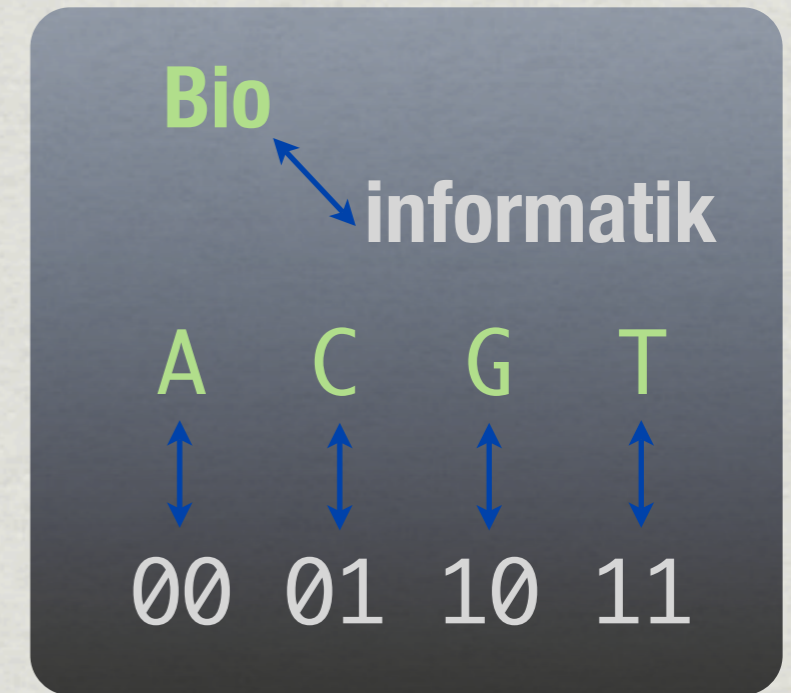


transtec



# Bioinformatik

- ▶ Probleme aus den Lebenswissenschaften
- ▶ Große Datenmengen, die von Menschen nicht analysiert werden können
- ▶ Lösung mittels theoretischer Methoden, Entwickeln von Algorithmen
- ▶ Selbst entwickelte Computerprogramme werden von Bioinformatikern online zur Verfügung gestellt



# Webdienste

- ▶ Isolierte Webseiten
- ▶ Oft einzelner „Programmierer“ (Doktorand)
- ▶ Wenig Anerkennung für gut gemachte Seiten
- ▶ Keine Langzeit-Betreuung

GLOBAL parameters

\*Unit Cell a  b  c   
alpha  beta  gamma

\*Data Resolution Low  High

\* Space Group

Fields with a \* (spacegroup, unit cell and data resolution) must be filled only if you provide an ASCII file. Otherwise, data are taken from MTZ header.

[reflection file in ASCII format \(H,K,L,F,sigF\)](#)

OR [reflection file in MTZ format \(truncate\)](#)

Email

Use Protein A    molA

Use Protein B    molB

AMoRe ( uncheck to skip )

[DATO.I3](#)

xyz	-1								
RATE	1	LMINs	6	LMINI	6	CUTR	0.0	STEP	2
TKEY 1	0-1	NUMR 1	25	CUTT 1	0.0	NPIC 1	25		
TKEY n	0-1	NUMR n	25	CUTT n	0.0	NPIC n	25		
NUMT	25	NITE	20						
CUTD (A)	5.0								

For more details about how to use dato.i3 parameters, please take a look at [AMoRe tutorial](#) ( pp 13-14 )

Resolution Range Low  High

[structure\(s\) to test with AMoRe](#)  
(i.e. just used in a simple AMoRe step, will be ignored by T-COFFEE and MODELLER)

PDB A test 1  
Code  chain   
or File

PDB B test 1  
Code  chain   
or File

# Motivation des Projekts

- ▶ Ziele für Bioinformatiker
  - ▶ Leicht erstellbar, externe Betreuung
  - ▶ Häufige Benutzung durch Wissenschaftler
- ▶ Ziele für Lebenswissenschaftler
  - ▶ Erreichbarkeit, Anleitung, einfache Bedienung
  - ▶ Integration verschiedener Programme
  - ▶ Arbeitsabläufe und Ergebnisse speichern

# Bestandsaufnahme

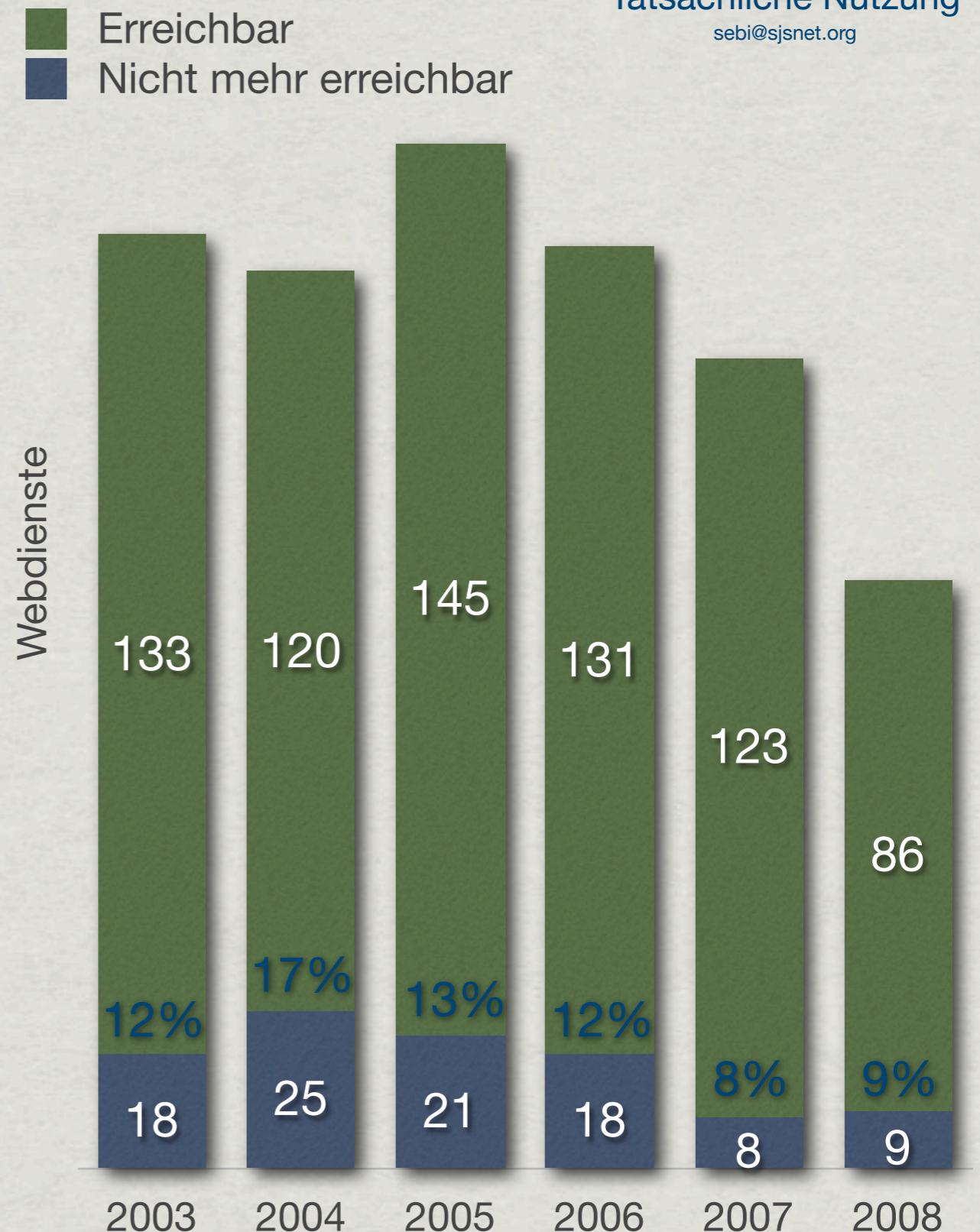
- ▶ Webdienste in der Ausgabe „Web Server Issue“ von *Nucleic Acids Research*
  - ▶ 6 Sonderausgaben seit 2003, 811 Webdienste
- ▶ Zusätzlich: [bioinformatics.ca/links\\_directory](http://bioinformatics.ca/links_directory), Publikationen in *Bioinformatics* und *BMC Bioinformatics*: ca. 2000 Webdienste



# NAR Web Server Issue

S. J. Schultheiß (2008)  
Bioinformatik Webdienste  
Tatsächliche Nutzung  
sebi@sjsnet.org

- ▶ Bereits viele Seiten nicht mehr erreichbar
- ▶ Statistische Untersuchung z.B. nach Land, Zeitschrift, Institution, Datum, ...
- ▶ Nach Nutzerstatistik fragen, auswerten



# Bisherige Lösungen

- ▶ Große Dienstleister wie *EBI*, *NCBI* bündeln Dienste
- ▶ Meta-Seiten und -Programme bieten Zugang zu verteilten Diensten (Grid): *CoMPAS Pro*, *Taverna*, *DAS*, *BioMart*
- ▶ Einheitliche Benutzeroberfläche für Kommandozeilentools: *Galaxy*

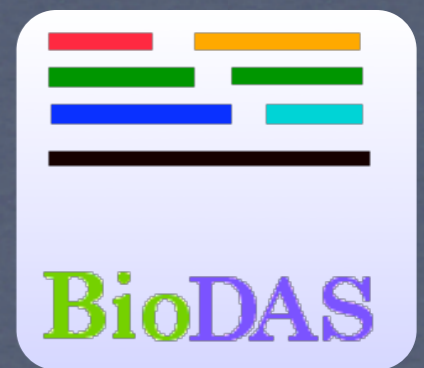
S.  
Bio  
Ta



CoMPAS Pro

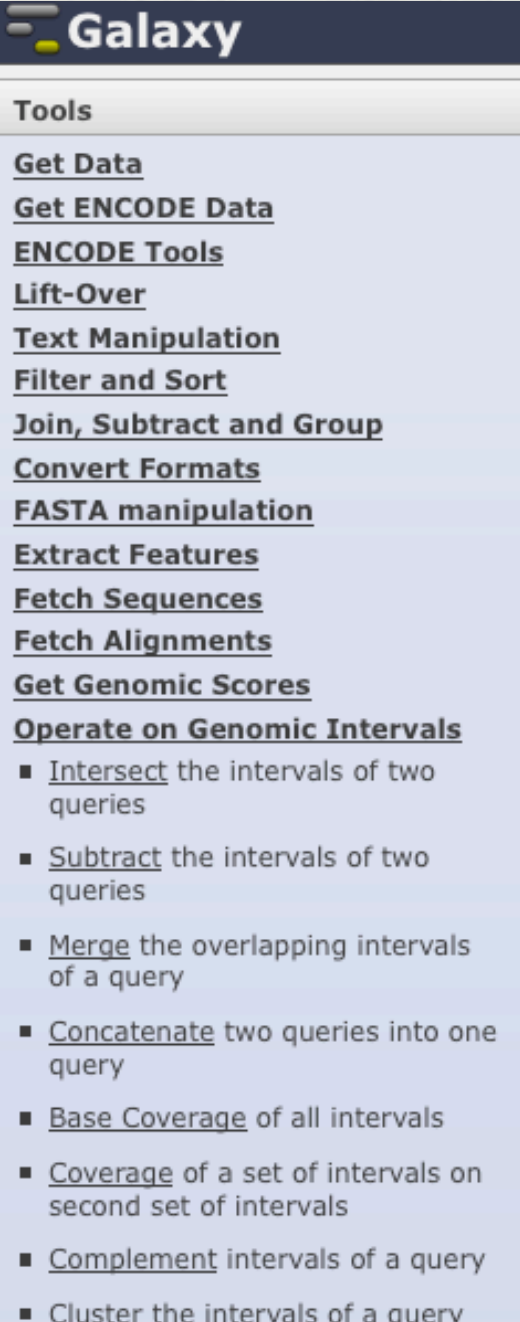


Taverna



# Galaxy

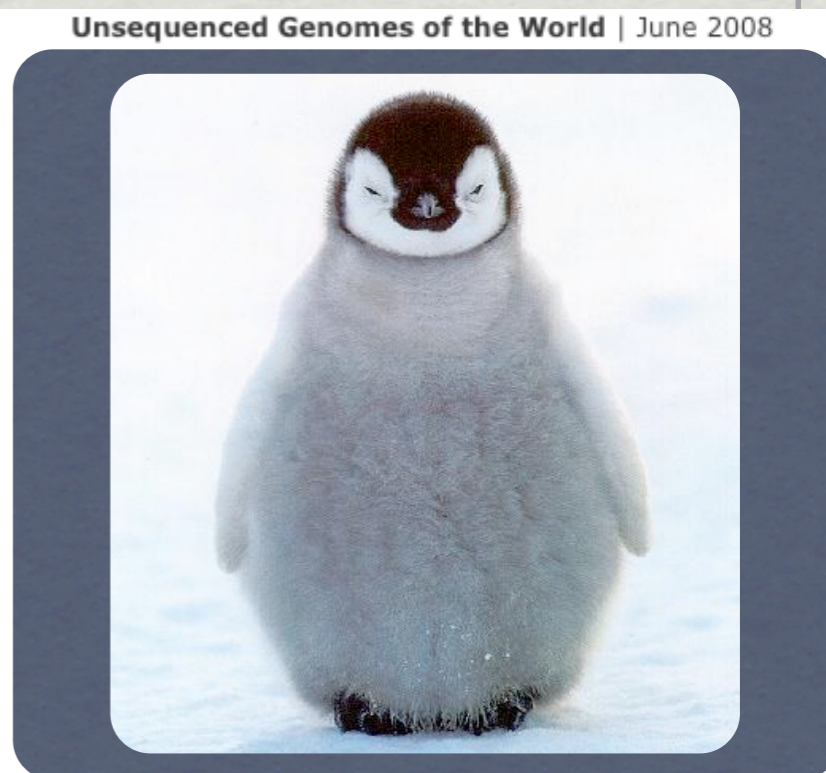
- ▶ Rahmenprogramm mit Standardtools, Integration eigener Programme (Open Source), erweiterbar
- ▶ Einfaches Verwalten von Experimentdaten, Weitergabe der Ergebnisse an andere Programme
- ▶ Entwicklerteam (Penn State und NYU, USA)
- ▶ Unser Treffen: Standarddefinition, offene Schnittstelle für Entwickler, alternative Oberfläche, Nachhaltigkeit



**Galaxy**

Tools

- [Get Data](#)
- [Get ENCODE Data](#)
- [ENCODE Tools](#)
- [Lift-Over](#)
- [Text Manipulation](#)
- [Filter and Sort](#)
- [Join, Subtract and Group](#)
- [Convert Formats](#)
- [FASTA manipulation](#)
- [Extract Features](#)
- [Fetch Sequences](#)
- [Fetch Alignments](#)
- [Get Genomic Scores](#)
- [Operate on Genomic Intervals](#)
  - [Intersect](#) the intervals of two queries
  - [Subtract](#) the intervals of two queries
  - [Merge](#) the overlapping intervals of a query
  - [Concatenate](#) two queries into one query
  - [Base Coverage](#) of all intervals
  - [Coverage](#) of a set of intervals on second set of intervals
  - [Complement](#) intervals of a query
  - [Cluster](#) the intervals of a query



### Galaxy is for Biologists

Use this site to access popular sources of data like the UCSC Table Browser. Run analyses right on the spot using a variety of integrated tools. Your results are always available and can be easily shared with others. Just [watch how](#).

### Galaxy is for Developers

Galaxy is an easy-to-use, open-source, scalable framework for tool and data integration. Stop wasting time writing interfaces and get your tools used by biologists! Galaxy includes everything you need to get started, so [download](#) and start [integrating!](#)



# Projektüberblick

- ▶ Internetadressen aus Publikationen extrahieren
- ▶ Datenbank der Internetadressen erstellen
- ▶ Wissenschaftliche Hilfskräfte füllen Datenbank
- ▶ Kontaktieren der Betreiber: Nutzerstatistiken, Informationen über Wartung, Aktualisierungen
- ▶ Auswertung, Abschlussbericht, Datenbank veröffentlichen, wissenschaftliche Publikation

2 w

1 w

3 m

1,5 m

# Danksagungen

- ▶ **TL-Stiftung und Kuratorium, Uni Tübingen**
- ▶ **Dr. Gunnar Rätsch (FML), Prof. Dr. Jan Lohmann (Uni HD)**
- ▶ **Galaxy Entwicklerteam**
- ▶ **Chris Malisi (CoMPAS Pro)**
- ▶ **Familie und Freunde**



EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



**transtec**



# Vielen Dank für Ihre Aufmerksamkeit!



EBERHARD KARLS  
UNIVERSITÄT  
TÜBINGEN



**transtec**

