

# Studienrichtung *Bioinformatik*

Prof. Burkhard Morgenstern

Abt. Bioinformatik  
Institut für Mikrobiologie und Genetik (IMG)

January 14, 2020



# Bioinformatik - was ist das?



The screenshot shows the homepage of the Fachgruppe Bioinformatik (FaBI). The header features the FaBI logo (a computer monitor with a DNA helix) and the text 'FaBI Fachgruppe Bioinformatik'. To the right is the logo of the Gesellschaft für Bioinformatik (GfBI) and a search bar. The navigation menu includes 'Home', 'Fachgruppe', 'Bioinformatik', 'Bioinformatik in Deutschland', 'News', 'Veranstaltungen', 'Stellenangebote', and 'Login'. The main content area has three featured articles: 'Was ist Bioinformatik?', 'Forschung in Deutschland', and 'Studium und Lehre'. Below these is a 'Herzlich willkommen' message and a 'Highlight' section for 'GCB2019'.

**FaBI** Fachgruppe Bioinformatik

Home Fachgruppe Bioinformatik Bioinformatik in Deutschland News Veranstaltungen Stellenangebote Login

**Was ist Bioinformatik?**  
Die Bioinformatik vereint Menschen aus der Informatik und den Naturwissenschaften aus den Lebenswissenschaften und hat sich als verbindende Brücke etabliert.  
[Read More](#)

**Forschung in Deutschland**  
Aufgrund von zunehmend steigenden Forschungsansätzen in den Lebenswissenschaften und der Verfügbarkeit von großen Hochdurchsatz-Datensätzen ist die Bioinformatik zu einer zentralen Disziplin geworden.  
[Read More](#)

**Studium und Lehre**  
Die Naturwissenschaften sind ein integraler Bestandteil der Bioinformatik. In den Lebenswissenschaften ist die Bioinformatik ein zentraler Bestandteil der Ausbildung.  
[Read More](#)

Herzlich willkommen auf der Webseite der Fachgruppe Bioinformatik (FaBI)

Im September 2014 haben sich die Bioinformatiker in Deutschland in der gemeinsamen Fachgruppe Bioinformatik (FaBI) vereint. Sie ging aus den entsprechenden Fachgruppen der vier Fachgesellschaften: GI (Gesellschaft für Informatik e.V.), DECHEMA (Gesellschaft für Chemische Technik und Biotechnologie e.V.), GBM (Gesellschaft für Biochemie und Molekularbiologie e.V.) und GDCh (Gesellschaft Deutscher Chemiker e.V.) hervor. Im Herbst 2015 trat die Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. (GMDS) bei. Die FaBI vertritt zurzeit über 750 Mitglieder und versteht sich als die gemeinsame Interessenvertretung der Bioinformatik-Forschung in Deutschland und als Ansprechpartnerin für Politik, Wirtschaft und Gesellschaft mit dem Ziel einer starken Informatik- und Bioinformatik-Forschung.

**Highlight**

**GCB2019**  
Die nächste GCB wird in München stattfinden. Das lokale Organisationskomitee, geleitet von Christoph Dieterich (akt. vom 17.-19. September 2018), ist Bioinformatiker, aber...  
[Read More](#)

Die *Fachgruppe Bioinformatik (FaBI)*

<http://bioinformatik.de/>



# Bioinformatik - was ist das?

Die Fachgruppe für Bioinformatik (FaBI) hat folgende Definition von Bioinformatik als Grundlage Ihrer Arbeit definiert:

*"Die Bioinformatik ist eine interdisziplinäre Wissenschaft. Unter Bioinformatik verstehen wir die Erforschung, Entwicklung und Anwendung computergestützter Methoden zur Beantwortung molekularbiologischer und biomedizinischer Fragestellungen. Im Fokus stehen Modelle und Algorithmen für Daten auf molekularer und zellbiologischer Ebene, beispielsweise für*

- *Genome und Gene,*
- *Gen- und Proteinexpression und -regulation,*
- *metabolische und regulatorische Pfade und Netzwerke,*
- *Strukturen von Biomakromolekülen, insb. DNA, RNA und Proteine,*
- *molekulare Interaktionen zwischen Biomakromolekülen untereinander und zwischen Biomakromolekülen und weiteren Substanzen wie beispielsweise Substraten, Transmittern, Botenstoffen und Inhibitoren sowie*
- *die molekulare Charakterisierung von Ökosystemen."*

Definition der *Bioinformatik* durch die *FaBI*

<http://bioinformatik.de/>



# Bioinformatik - was ist das?

Die Fachgruppe für Bioinformatik (FaBI) hat folgende Definition von Bioinformatik als Grundlage Ihrer Arbeit definiert:

*"Die Bioinformatik ist eine interdisziplinäre Wissenschaft. Unter Bioinformatik verstehen wir die Erforschung, Entwicklung und Anwendung computergestützter Methoden zur Beantwortung molekularbiologischer und biomedizinischer Fragestellungen. Im Fokus stehen Modelle und Algorithmen für Daten auf molekularer und zellbiologischer Ebene, beispielsweise für*

- *Genome und Gene,*
- *Gen- und Proteinexpression und -regulation,*
- *metabolische und regulatorische Pfade und Netzwerke,*
- *Strukturen von Biomakromolekülen, insb. DNA, RNA und Proteine,*
- *molekulare Interaktionen zwischen Biomakromolekülen untereinander und zwischen Biomakromolekülen und weiteren Substanzen wie beispielsweise Substraten, Transmittern, Botenstoffen und Inhibitoren sowie*
- *die molekulare Charakterisierung von Ökosystemen."*

Definition der *Bioinformatik* durch die *FaBI*

<http://bioinformatik.de/>



# Bioinformatik - was ist das?

Die Fachgruppe für Bioinformatik (FaBI) hat folgende Definition von Bioinformatik als Grundlage Ihrer Arbeit definiert:

*"Die Bioinformatik ist eine interdisziplinäre Wissenschaft. Unter Bioinformatik verstehen wir die Erforschung, Entwicklung und Anwendung computergestützter Methoden zur Beantwortung molekularbiologischer und biomedizinischer Fragestellungen. Im Fokus stehen Modelle und Algorithmen für Daten auf molekularer und zellbiologischer Ebene, beispielsweise für*

- Genome und Gene,
- Gen- und Proteinexpression und -regulation,
- metabolische und regulatorische Pfade und Netzwerke,
- Strukturen von Biomakromolekülen, insb. DNA, RNA und Proteine,
- molekulare Interaktionen zwischen Biomakromolekülen untereinander und zwischen Biomakromolekülen und weiteren Substanzen wie beispielsweise Substraten, Transmittern, Botenstoffen und Inhibitoren sowie
- die molekulare Charakterisierung von Ökosystemen."

Definition der *Bioinformatik* durch die *FaBI*

<http://bioinformatik.de/>



Arbeitsfelder:

- Grundlagenforschung:  
Algorithmen, Modelle, Software
- Angewandte Bioinformatik:  
Biologische Datenanalyse
- Datenbanken, Infrastruktur



Arbeitsmöglichkeiten:

- Universitäten, Forschungsinstitute
- Bioinformatik- und Biotechfirmen
- Pharmaindustrie



Arbeitsmöglichkeiten:

- Universitäten, Forschungsinstitute
- Bioinformatik- und Biotechfirmen
- Pharmaindustrie

Job-Aussichten: Ausgezeichnet!





# Bioinformatik-AGs in Göttingen

- Theoretische Informatik (Institut für Informatik), Prof. S. Waack
- Bioinformatik (Biologische Fakultät), Prof. B. Morgenstern
- Biostatistik/Bioinformatik (UMG), Prof. T. Beißbarth
- Transkriptomanalyselabor (UMG), Dr. G. Salinas
- MPI für Biophysikalische Chemie, Dr. J. Söding
- Züchtungsinformatik (Agrarwissenschaften), Prof. A. Schmitt
- Angewandte Bioinformatik (Biologische Fakultät), Prof. J. de Vries

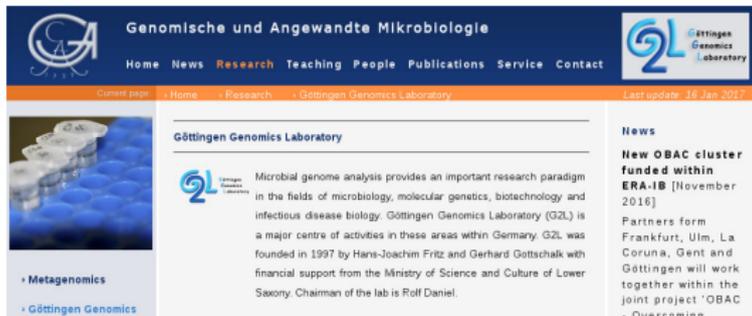


# Bioinformatik-AGs in Göttingen

Außerdem: Stark wachsender Bedarf an Bioinformatikern in

- Molekularbiologischen AGs
- Core Facilities

Z.B.: Göttingen Genome Laboratory (G<sub>2</sub>L):



The screenshot shows the website for the Göttingen Genomics Laboratory (G<sub>2</sub>L). The header features the logo of the Institute of Genomics and Applied Microbiology (GA) and the G<sub>2</sub>L logo. Navigation links include Home, News, Research, Teaching, People, Publications, Service, and Contact. A breadcrumb trail shows the current page is 'Göttingen Genomics Laboratory'. The main content area is titled 'Göttingen Genomics Laboratory' and includes a description of the lab's focus on microbial genome analysis. A 'News' section highlights the funding of the OBAC cluster within the ERA-IB project in November 2016. A sidebar on the left lists 'Metagenomics' and 'Göttingen Genomics' as categories.

Genomische und Angewandte Mikrobiologie

Home News **Research** Teaching People Publications Service Contact

Current page: Home Research Göttingen Genomics Laboratory Last update: 16 Jan 2017

**Göttingen Genomics Laboratory**

Microbial genome analysis provides an important research paradigm in the fields of microbiology, molecular genetics, biotechnology and infectious disease biology. Göttingen Genomics Laboratory (G<sub>2</sub>L) is a major centre of activities in these areas within Germany. G<sub>2</sub>L was founded in 1997 by Hans-Joachim Fritz and Gerhard Gottschalk with financial support from the Ministry of Science and Culture of Lower Saxony. Chairman of the lab is Rolf Daniel.

**News**

**New OBAC cluster funded within ERA-IB** [November 2016]

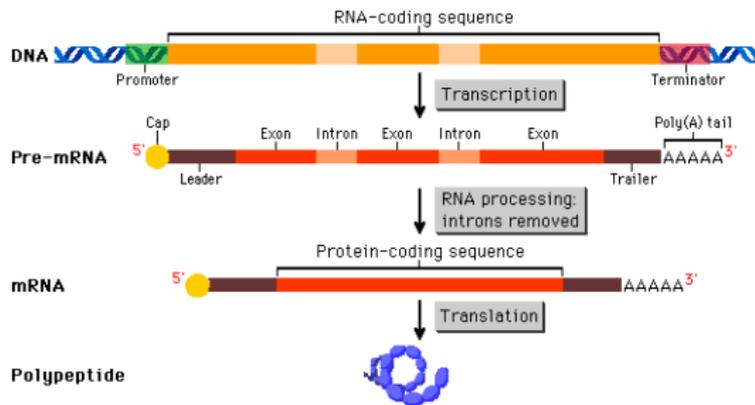
Partners from Frankfurt, Ulm, La Coruna, Gent and Göttingen will work together within the joint project 'OBAC - Overcoming

• Metagenomics

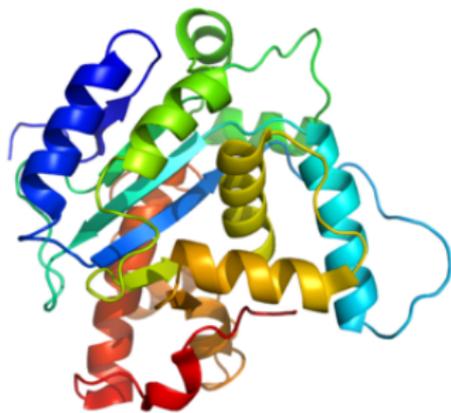
• Göttingen Genomics







Übersetzung *DNA* → *mRNA* → *Protein*



Protein-Struktur

```
seq1    W T Y I V M R E A Q Y E S A Q
seq2    R C L V M R E A Q E W
seq3    Y I M R D V Q Q E R A
seq4    A L W I A M R E V Q Y E S A
```

Arbeitsschwerpunkt:

Algorithmen zum Vergleich von DNA- und Protein-Sequenzen

(A) Sequenz-Alignment



seq1	W	T	Y	I	V	M	R	E	A	Q	Y	E	S	A	Q
seq2	-	R	C	L	V	M	R	E	A	Q	-	E	W	-	-
seq3	-	-	Y	I	-	M	R	D	V	Q	Q	E	R	A	-
seq4	A	L	W	I	A	M	R	E	V	Q	Y	E	S	A	-

Arbeitsschwerpunkt:

Algorithmen zum Vergleich von DNA- und Protein-Sequenzen

(A) Sequenz-Alignment



seq1	W	T	Y	I	V	M	R	E	A	Q	Y	E	S	A	Q
seq2	-	R	C	L	V	M	R	E	A	Q	-	E	W	-	-
seq3	-	-	Y	I	-	M	R	D	V	Q	Q	E	R	A	-
seq4	A	L	W	I	A	M	R	E	V	Q	Y	E	S	A	-

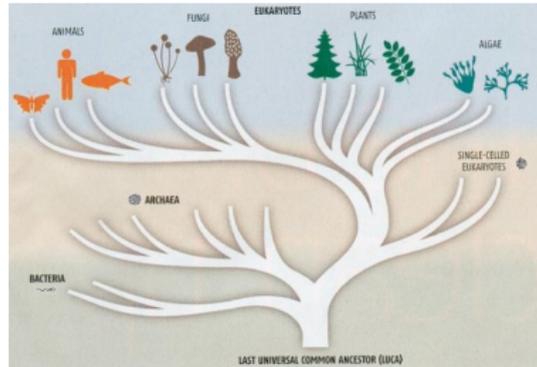
Arbeitsschwerpunkt:

Algorithmen zum Vergleich von DNA- und Protein-Sequenzen

(A) Sequenz-Alignment

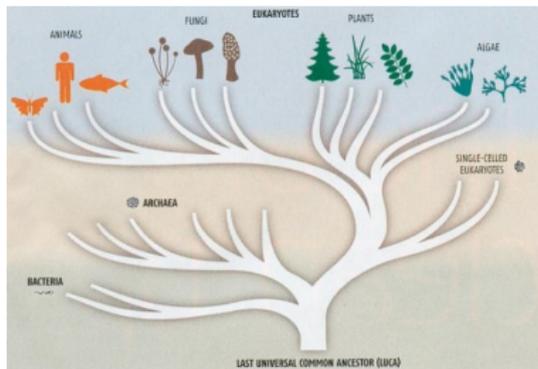


## Rekonstruktion phylogenetischer Bäume:



<http://universe-review.ca>

## Rekonstruktion phylogenetischer Bäume:

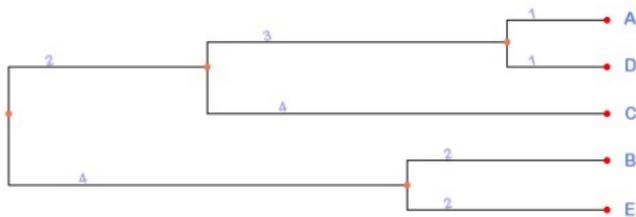


<http://universe-review.ca>

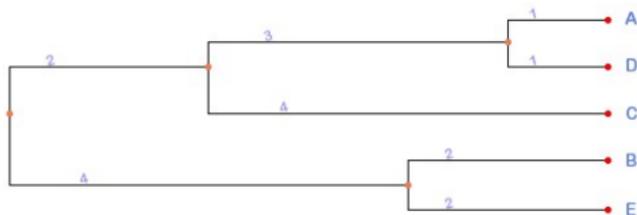
Ansatz: Für Matrix mit paarweisen *Distanzen* zwischen Objekten, finde passenden Baum mit Objekten an den *Blättern*



	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>E</i>
<i>A</i>	0	12	8	2	12
<i>B</i>		0	12	12	4
<i>C</i>			0	8	12
<i>D</i>				0	12
<i>E</i>					0



	A	B	C	D	E
A	0	12	8	2	12
B		0	12	12	4
C			0	8	12
D				0	12
E					0



Distanzen aus *Alignments* geschätzt.

Problem: Berechnung von *Alignments* zeitaufwändig!



## (B) Alignment-freier Sequenz-Vergleich



## (B) Alignment-freier Sequenz-Vergleich

- Für festes pattern  $P$ , z.B. '11010001'  
(‘Match Positionen’ und ‘*Don't-care*’ Positionen)



## (B) Alignment-freier Sequenz-Vergleich

- Für festes pattern  $P$ , z.B. '11010001'  
(‘Match Positionen’ und ‘*Don't-care*’ Positionen)
- Finde alle *Spaced-Word-Matche* bzgl.  $P$



## (B) Alignment-freier Sequenz-Vergleich

- Für festes pattern  $P$ , z.B. '11010001'  
(‘Match Positionen’ und ‘Don’t-care’ Positionen)
- Finde alle *Spaced-Word-Matche* bzgl.  $P$

Example (Spaced-Word-Match bzgl. ' $P = 11010001$ ')

							1	1	0	1	0	0	0	1			
$S_1$	D	W	Y	S	L	G	C	Y	M	S	R	C	I	R	V	M	Q
$S_2$	R	W	V	C	Y	L	S	K	C	L	R	C	A	G	C		

## (B) Alignment-freier Sequenz-Vergleich

- Für festes pattern  $P$ , z.B. '11010001'  
(‘Match Positionen’ und ‘Don’t-care’ Positionen)
- Finde alle *Spaced-Word-Matche* bzgl.  $P$

### Example (Spaced-Word-Match bzgl. ' $P = 11010001$ ')

							1	1	0	1	0	0	0	1			
$S_1$	D	W	Y	S	L	G	C	Y	M	S	R	C	I	R	V	M	Q
$S_2$	R	W	V	C	Y	L	S	K	C	L	R	C	A	G	C		
				1	1	0	1	0	0	0	1						



## (B) Alignment-freier Sequenz-Vergleich

- Für festes pattern  $P$ , z.B. '11010001'  
(‘Match Positionen’ und ‘Don’t-care’ Positionen)
- Finde alle *Spaced-Word-Matche* bzgl.  $P$

Example (Spaced-Word-Match bzgl. ‘ $P = 11010001$ ’)

$S_1$	...	C	Y	M	S	R	C	I	R	...
$S_2$	...	C	Y	L	S	K	C	L	R	...
		1	1	0	1	0	0	0	1	



## (B) Alignment-freier Sequenz-Vergleich

- Für festes pattern  $P$ , z.B. '11010001'  
(‘Match Positionen’ und ‘*Don't-care*’ Positionen)
- Finde alle *Spaced-Word-Matche* bzgl.  $P$

Example (Spaced-Word-Match bzgl. ' $P = 11010001$ ')

$S_1$	...	C	Y	M	S	R	C	I	R	...
$S_2$	...	C	Y	L	S	K	C	L	R	...
		1	1	0	1	0	0	0	1	

Nehme *don't-care* Positionen, um Sequenzählichkeit zu messen



## (B) Alignment-freier Sequenz-Vergleich

- Für festes pattern  $P$ , z.B. '11010001'  
(‘Match Positionen’ und ‘*Don't-care*’ Positionen)
- Finde alle *Spaced-Word-Matche* bzgl.  $P$

Example (Spaced-Word-Match bzgl. ' $P = 11010001$ ')

$S_1$	...	$M$	$R$	$C$	$I$	...
$S_2$	...	$L$	$K$	$C$	$L$	...
		0	0	0	0	

Nehme *don't-care* Positionen, um Sequenzähnlichkeit zu messen



## (B) Alignment-freier Sequenz-Vergleich

- Für festes pattern  $P$ , z.B. '11010001'  
(‘Match Positionen’ und ‘*Don't-care*’ Positionen)
- Finde alle *Spaced-Word-Matche* bzgl.  $P$

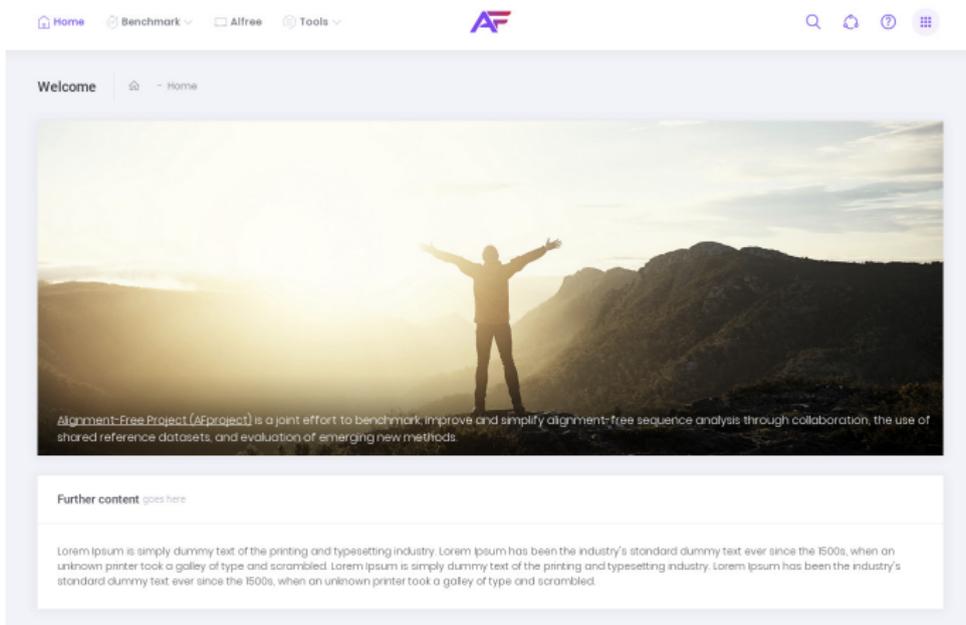
Example (Spaced-Word-Match bzgl. ' $P = 11010001$ ')

$S_1$	...	$M$	$R$	$C$	$I$	...
$S_2$	...	$L$	$K$	$C$	$L$	...
		0	0	0	0	

Nehme *don't-care* Positionen, um Sequenzähnlichkeit zu messen



## Internationale Kollaboration um Alignment-freie Methoden zu vergleichen / evaluieren



Home Benchmark Alfree Tools

AF

Welcome Home

Alignment-Free Project (AFProject) is a joint effort to benchmark, improve and simplify alignment-free sequence analysis through collaboration, the use of shared reference datasets, and evaluation of emerging new methods.

Further content goes here

Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled. Lorem Ipsum is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled.



## Performanz der evaluierten Methoden

AFKS--spearman	0.46	0.63	0.72	0.67	0.42	0.55		0.54	0.47	1.00	0.50	0.77
AFKS--squared_chord	0.51	0.71	0.77	0.67	0.46	0.36		0.43	0.29	1.00	0.50	0.72
alfpy--angle_cos_diss	0.54	0.64	0.72	0.66	0.46	0.18	0.95			1.00	0.50	0.93
alfpy--angle_cos_evolution	0.54	0.64	0.72	0.66	0.46	0.18	0.95			1.00	0.50	0.93
alfpy--braycurtis	0.59	0.73	0.78	0.67	0.50	0.55	0.91			1.00	0.50	0.93
alfpy--canberra	0.74	0.70	0.76	0.68	0.46	0.55	0.91			1.00	0.50	0.93
alfpy--chebyshev	0.68	0.65	0.72	0.30	0.27	0.09	0.36			0.20	0.38	0.35
alfpy--euclid_norm	0.57	0.66	0.73	0.66	0.42	0.27	0.95			0.80	0.50	0.90
alfpy--euclid_squared	0.57	0.66	0.73	0.66	0.50	0.27	0.95			1.00	0.50	0.93
alfpy--google	0.59	0.74	0.78	0.66	0.50	0.55	0.91			1.00	0.50	0.91
alfpy--jpd	0.63	0.63	0.72	0.67	0.46	0.64	0.91			1.00	0.58	0.93
alfpy--kdd	0.57	0.63	0.71	0.65	0.31	0.09	0.91			0.80	0.29	0.63
alfpy--lcc	0.53	0.65	0.73	0.66	0.46	0.18	0.91			1.00	0.50	0.93
alfpy--manhattan	0.59	0.66	0.74	0.67	0.50	0.55	0.91			1.00	0.50	0.93
alfpy--minkowski	0.57	0.66	0.73	0.66	0.42	0.27	0.95			0.80	0.50	0.90
alfpy--wmetric		0.64	0.72	0.33								
ALFRED-G	0.52	0.59	0.69			0.54	0.91			0.80	0.54	0.91
andi					0.92	0.18	0.77	0.73		0.20	0.92	
cale--cvtree	0.53				0.42	0.73	0.41	0.56	0.57	1.00	0.50	0.94
cale--d2shepp	0.60				0.42	0.82	0.73	0.56	0.68	1.00	0.50	0.72
cale--d2star	0.60				0.46	0.73	0.68	0.41	0.66	1.00	0.50	0.91
co-phylog					0.88		0.32	0.79		0.20	0.92	0.68
FFP	0.50	0.63	0.70	0.67	0.77	0.82	0.91	0.57	0.57	0.20	0.79	0.93
FSWM / Read-SpaM					0.88	0.73	0.95	0.67	0.52	0.00	0.83	0.67
jD2Stat--D2n	0.51	0.64	0.72	0.69	0.46		0.82			0.40	0.46	0.03
jD2Stat--D2S	0.59	0.59	0.70	0.62	0.46		0.95			1.00	0.54	0.93
jD2Stat--D2SI	0.60	0.59	0.69	0.56	0.46		0.86			1.00	0.50	0.77
kmacs	0.46	0.64	0.71		0.50		0.95			0.80	0.46	0.89
kr	0.46				0.77					0.20	0.71	0.53
kSNP3					0.85	0.55	0.91	0.63	0.56	0.20	0.79	0.86
KWIP								0.43	0.51			
ltw-kernel	0.52	0.53	0.59	0.63	0.50	0.36	0.36			1.00	0.38	0.32
ltw-ncd	0.51	0.69	0.76	0.46	0.19		0.14			0.20	0.12	0.21
mash	0.53	0.68	0.75	0.60	0.85	0.82	0.95	0.66	0.60	0.20	0.88	0.92
multi-SpaM					0.73	0.91	0.82			0.40	0.83	0.64
phyloium					0.96	0.36	0.86			0.20	0.92	
rtD-phylogeny	0.61	0.58	0.64	0.47	0.19	0.09	0.95			0.80	0.25	0.09
spaced	0.52	0.60	0.70	0.67	0.65	0.55	0.95			0.20	0.67	0.89
UnderlyingApproach	0.53						0.86					

## Performanz der evaluierten Methoden

AFKS--spearman	0.46	0.63	0.72	0.67	0.42	0.55		0.54	0.47	1.00	0.50	0.77
AFKS--squared_chord	0.51	0.71	0.77	0.67	0.46	0.36		0.43	0.29	1.00	0.50	0.72
alfpy--angle_cos_diss	0.54	0.64	0.72	0.66	0.46	0.18	0.95			1.00	0.50	0.93
alfpy--angle_cos_evolution	0.54	0.64	0.72	0.66	0.46	0.18	0.95			1.00	0.50	0.93
alfpy--braycurtis	0.59	0.73	0.78	0.67	0.50	0.55	0.91			1.00	0.50	0.93
alfpy--canberra	0.74	0.70	0.76	0.68	0.46	0.55	0.91			1.00	0.50	0.93
alfpy--chebyshev	0.68	0.65	0.72	0.30	0.27	0.09	0.36			0.20	0.38	0.35
alfpy--euclid_norm	0.57	0.66	0.73	0.66	0.42	0.27	0.95			0.80	0.50	0.90
alfpy--euclid_squared	0.57	0.66	0.73	0.66	0.50	0.27	0.95			1.00	0.50	0.93
alfpy--google	0.59	0.74	0.78	0.66	0.50	0.55	0.91			1.00	0.50	0.91
alfpy--jpd	0.63	0.63	0.72	0.67	0.46	0.64	0.91			1.00	0.58	0.93
alfpy--kdd	0.57	0.63	0.71	0.65	0.31	0.09	0.91			0.80	0.29	0.63
alfpy--lcc	0.53	0.65	0.73	0.66	0.46	0.18	0.91			1.00	0.50	0.93
alfpy--manhattan	0.59	0.66	0.74	0.67	0.50	0.55	0.91			1.00	0.50	0.93
alfpy--minkowski	0.57	0.66	0.73	0.66	0.42	0.27	0.95			0.80	0.50	0.90
alfpy--wmetric		0.64	0.72	0.33								
ALFRED-G	0.52	0.59	0.69			0.54	0.91			0.80	0.54	0.91
andi					0.92	0.18	0.77	0.73		0.20	0.92	
cale--cvtree	0.53				0.42	0.73	0.41	0.56	0.57	1.00	0.50	0.94
cale--d2shepp	0.60				0.42	0.82	0.73	0.56	0.68	1.00	0.50	0.72
cale--d2star	0.60				0.46	0.73	0.68	0.41	0.66	1.00	0.50	0.91
co-phylog					0.88		0.32	0.79		0.20	0.92	0.68
FFP	0.50	0.63	0.70	0.67	0.77	0.82	0.91	0.57	0.57	0.20	0.79	0.93
<b>FSWM / Read-SpaM</b>					0.88	0.73	0.95	0.67	0.52	0.00	0.83	0.67
jD2Stat--D2n	0.51	0.64	0.72	0.69	0.46		0.82			0.40	0.46	0.03
jD2Stat--D2S	0.59	0.59	0.70	0.62	0.46		0.95			1.00	0.54	0.93
jD2Stat--D2SI	0.60	0.59	0.69	0.56	0.46		0.86			1.00	0.50	0.77
<b>kmacs</b>	0.46	0.64	0.71		0.50		0.95			0.80	0.46	0.89
kr	0.46				0.77					0.20	0.71	0.53
kSNP3					0.85	0.55	0.91	0.63	0.56	0.20	0.79	0.86
KWIP								0.43	0.51			
lzw-kernel	0.52	0.53	0.59	0.63	0.50	0.36	0.36			1.00	0.38	0.32
lzw-nod	0.51	0.69	0.76	0.46	0.19		0.14			0.20	0.12	0.21
mash	0.53	0.68	0.75	0.60	0.85	0.82	0.95	0.66	0.60	0.20	0.88	0.92
<b>multi-SpaM</b>					0.73	0.91	0.82			0.40	0.83	0.64
phyloium					0.96	0.36	0.86			0.20	0.92	
rtD-phylogeny	0.61	0.58	0.64	0.47	0.19	0.09	0.95			0.80	0.25	0.09
<b>spaced</b>	0.52	0.60	0.70	0.67	0.65	0.55	0.95			0.20	0.67	0.89
UnderlyingApproach	0.53						0.86					

## **Fast alignment-free sequence comparison using spaced-word frequencies**

Chris-Andre Leimeister<sup>1,\*</sup>, Marcus Boden<sup>1</sup>, Sebastian Horwege<sup>1</sup>, Sebastian Lindner<sup>1</sup> and Burkhard Morgenstern<sup>1,2</sup>

<sup>1</sup>Department of Bioinformatics, University of Göttingen, Institute of Microbiology and Genetics, 37073 Göttingen, Germany and <sup>2</sup>Université d'Évry Val d'Essonne, Laboratoire Statistique et Génome, UMR CNRS 8071, USC INRA, 91037 Évry, France

Associate Editor: John Hancock

Publikationen von MSc- und BSc-Studierenden (Info oder Bio)



## **Fast alignment-free sequence comparison using spaced-word frequencies**

Chris-Andre Leimeister<sup>1,\*</sup>, Marcus Boden<sup>1</sup>, Sebastian Horwege<sup>1</sup>, Sebastian Lindner<sup>1</sup> and Burkhard Morgenstern<sup>1,2</sup>

<sup>1</sup>Department of Bioinformatics, University of Göttingen, Institute of Microbiology and Genetics, 37073 Göttingen, Germany and <sup>2</sup>Université d'Évry Val d'Essonne, Laboratoire Statistique et Génome, UMR CNRS 8071, USC INRA, 91037 Évry, France

Associate Editor: John Hancock

Publikationen von MSc- und BSc-Studierenden (Info oder Bio)



Published online 14 May 2014

*Nucleic Acids Research*, 2014, Vol. 42, Web Server issue W7–W11  
doi: 10.1093/nar/gku398

## ***Spaced words* and *kmacs*: fast alignment-free sequence comparison based on inexact word matches**

**Sebastian Horwege<sup>1,\*</sup>, Sebastian Lindner<sup>1</sup>, Marcus Boden<sup>1</sup>**, Klas Hatje<sup>2</sup>, Martin Kollmar<sup>2</sup>,  
**Chris-André Leimeister<sup>1</sup>** and Burkhard Morgenstern<sup>1,3</sup>

<sup>1</sup>University of Göttingen, Institute of Microbiology and Genetics, Department of Bioinformatics, Goldschmidtstraße 1, 37073 Göttingen, Germany, <sup>2</sup>Max-Planck-Institute for Biophysical Chemistry, Department of NMR-based Structural Biology, Group Systems Biology of Motor Proteins, Am Fassberg 11, 37077 Göttingen, Germany and <sup>3</sup>Université d'Évry Val d'Essonne, Laboratoire Statistique et Génome, UMR CNRS 8071, USC INRA, 23 Boulevard de France, 91037 Évry, France

Publikationen von MSc- und BSc-Studierenden (Info oder Bio)



RESEARCH ARTICLE

## *rasbhari*: Optimizing Spaced Seeds for Database Searching, Read Mapping and Alignment-Free Sequence Comparison

Lars Hahn<sup>1\*</sup>, Chris-André Leimeister<sup>1</sup>, Rachid Ounit<sup>2</sup>, Stefano Lonardi<sup>2</sup>,  
Burkhard Morgenstern<sup>1,3</sup>

Publikationen von MSc- und BSc-Studierenden (Info oder Bio)



Morgenstern et al. *Algorithms Mol Biol* (2017) 12:27  
<https://doi.org/10.1186/s13015-017-0118-8>

Algorithms for  
Molecular Biology

RESEARCH

Open Access

## Phylogeny reconstruction based on the length distribution of $k$ -mismatch common substrings



Burkhard Morgenstern<sup>\*</sup>, **Svenja Schöbel** and Chris-André Leimeister

Publikationen von MSc- und BSc-Studierenden (Info oder Bio)





GigaScience, 8, 2019, 1–14

doi: 10.1093/gigascience/giy148

Advance Access Publication Date: 7 December 2018

Research

RESEARCH

## *Prot-SpaM*: fast alignment-free phylogeny reconstruction based on whole-proteome sequences

Chris-Andre Leimeister <sup>1,\*</sup>, Jendrik Schellhorn <sup>1,†</sup>, Svenja Dörner <sup>1</sup>, Michael Gerth <sup>2</sup>, Christoph Bleidorn <sup>3,4</sup> and Burkhard Morgenstern <sup>1,5</sup>

Publikationen von MSc- und BSc-Studierenden (Info oder Bio)



Zielezinski *et al. Genome Biology* (2019) 20:144  
<https://doi.org/10.1186/s13059-019-1755-7>

Genome Biology

RESEARCH

Open Access

## Benchmarking of alignment-free sequence comparison methods

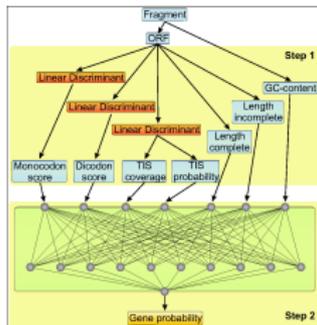


Andrzej Zielezinski<sup>1</sup>, Hani Z. Girgis<sup>2</sup>, Guillaume Bernard<sup>3</sup>, Chris-Andre Leimeister<sup>4</sup>, Kujin Tang<sup>5</sup>, Thomas Dencker<sup>4</sup>, Anna Katharina Lau<sup>4</sup>, Sophie Röhring<sup>4</sup>, Jae Jin Choi<sup>6,7</sup>, Michael S. Waterman<sup>5,8</sup>, Matteo Comin<sup>9</sup>, Sung-Hou Kim<sup>6,7</sup>, Susana Vinga<sup>10,11</sup>, Jonas S. Almeida<sup>12</sup>, Cheong Xin Chan<sup>13</sup>, Benjamin T. James<sup>2</sup>, Fengzhu Sun<sup>5,8</sup>, Burkhard Morgenstern<sup>4</sup> and Wojciech M. Karlowski<sup>1\*</sup> 

Publikationen von MSc- und BSc-Studierenden (Info oder Bio)



## Arbeitsschwerpunkt: Maschinelles Lernen in der Bioinformatik (Dr. Peter Meinicke)



# Studienplan

	Biologie		Bioinformatik	
<b>1. Semester</b>	—		—	
<b>2. Semester</b>	B.Bio-NF.102 Ringvorlesung II (VL)	8 C / 6 SWS	—	
<b>3. Semester</b>	B.Bio-NF.117.Mp Genomanalyse	6 C / 4 SWS	B.Inf.1501 Algorithmen der Bioinformatik I (VL + UE)	5 C / 4 SWS
	<i>optional: B.Bio-NF.119.1 Kognitive Neurowissenschaften (VL) alternativ im 5. Semester</i>	(3 C / 2 SWS)	B.Inf.1502 Molekularbiologische Datenbanken (VL)	5 C / 3 SWS
	<i>optional: B.Bio-NF.119.4 Biologische Psychologie I (VL) alternativ im 5. Semester</i>	(4 C / 2 SWS)		
<b>4. Semester</b>	B.Bio-NF.129 Genetik und mikrobielle Zellbiologie	6 C / 4 SWS	<i>optional: B.Inf.1503: Proseminar Bioinformatik (kann ab dem 3. Sem. belegt werden, wird jedes Semester angeboten)</i>	5 C / 2 SWS
	<i>optional: B.Bio-NF.119.2 Theoretische Neurowissenschaften (VL) alternativ im 6. Semester</i>	(4 C / 3 SWS)	<i>optional: B.Inf.1811 Vertiefte Angewandte Informatik im forschungsbezogenem Praktikum (kann ab dem 3. Sem. belegt werden, wird jedes Semester angeboten)</i>	(10 C / 1 SWS)
	<i>optional: B.Bio-NF.119.3 Neuro- und Verhaltensbiologie (VL) alternativ im 6. Semester</i>	(3 C / 2 SWS)		
<b>5. Semester</b>	<i>optional: SK.Bio.305 Biostatistik mit R (UE) Block in der vorlesungsfreien Zeit</i>	(3 C / 2 SWS)	B.Inf.1504 Maschinelles Lernen in der Bioinformatik (VL + UE)	5C / 4 SWS
<b>6. Semester</b>	—		<i>optional: B.Inf.1810 Angewandte Informatik im forschungsbezogenem Praktikum (kann ab dem 3. Sem. belegt werden, wird jedes Semester angeboten)</i>	(5 C / 0.5 SWS)
	<b>20-34 C</b>			<b>20-35 C</b>



BSc Projekt in Bioinformatik:



BSc Projekt in Bioinformatik:

- Implementierung oder Modifizierung neuer Ansätze



## BSc Projekt in Bioinformatik:

- Implementierung oder Modifizierung neuer Ansätze
- Anwendung auf biol. Daten, in Kooperationen



## BSc Projekt in Bioinformatik:

- Implementierung oder Modifizierung neuer Ansätze
- Anwendung auf biol. Daten, in Kooperationen
- Evaluieren von neu entwickelten Methoden



## BSc Projekt in Bioinformatik:

- Implementierung oder Modifizierung neuer Ansätze
- Anwendung auf biol. Daten, in Kooperationen
- Evaluieren von neu entwickelten Methoden
- Tiefe Biologie-Kenntnisse i.d.R. nicht erforderlich



BSc Projekt in Bioinformatik:

- Implementierung oder Modifizierung neuer Ansätze
- Anwendung auf biol. Daten, in Kooperationen
- Evaluieren von neu entwickelten Methoden
- Tiefe Biologie-Kenntnisse i.d.R. nicht erforderlich

Schwerpunkt in Bioinformatik auch möglich *ohne* Anwendungsfach Biologie ('Anwendungsorientierte Systementwicklung')



## BSc Projekt in Bioinformatik:

- Implementierung oder Modifizierung neuer Ansätze
- Anwendung auf biol. Daten, in Kooperationen
- Evaluieren von neu entwickelten Methoden
- Tiefe Biologie-Kenntnisse i.d.R. nicht erforderlich

Schwerpunkt in Bioinformatik auch möglich *ohne* Anwendungsfach Biologie ('Anwendungsorientierte Systementwicklung')



Für BSc Projekt in Bioinformatik in Abt. Morgenstern:



Für BSc Projekt in Bioinformatik in Abt. Morgenstern:

- Voraussetzung: Modul *Algorithmen der Bioinformatik I*



Für BSc Projekt in Bioinformatik in Abt. Morgenstern:

- Voraussetzung: Modul *Algorithmen der Bioinformatik I*
- Wünschenswert: Modul *Genomanalyse*



*Kontakt:*

**Prof. Burkhard Morgenstern**

`bmorgen@gwdg.de`

(Sequenzanalyse)

**Dr. Peter Meinicke**

`pmeinic@gwdg.de`

(Maschinelles Lernen)

