33rd Epilog

Celebrating Excellence

Distinguished Young Alumn Award
Siemens Awards for Excellence
Bachelor with Honors
Best Poster Award

TU WIEN Informatics

Welcome Address

- Sabine Seidler – Rector, TU Wien
- Gerti Kappel – Dean, TU Wien Informatics
- Host: Stefan Woltran – Vice Dean of Academic Affairs, TU Wien Informatics

33rd Epilog

Celebrating Excellence

Distinguished Young Alumn Award

Siemens Awards for Excellence

Bachelor with Honors

Best Poster Award

TU WIEN Informatics

Axel Brunnbauer

Model-Based Reinforcement Learning for Autonomous Racing

Distinguished Young Alumn Candidate

Axel Brunnbauer
Supervisor: Radu Grosu
191-01 Cyber-Physical Systems

Tobias Geibinger

Investigating Constraint Programming and Hybrid Answer-set Solving for Industrial Test Laboratory Scheduling

Distinguished Young Alumn Candidate

Tobias Geibinger
Supervisor: Nysret Musliu
192-02 Databases and Artificial Intelligence

Viktoriia Korchemna

The Complexity of Bayesian Network Learning: Revisiting the Superstructure

Distinguished Young Alumn Candidate

Viktoriia Korchemna
Supervisor: Robert Ganian
192-01 Algorithms and Complexity

Sebastian Mazza

Homomorphic-Encrypted Volume Rendering

Distinguished Young Alumn Candidate

Sebastian Mazza
Supervisor: Ivan Viola
193-02 Computer Graphics

33rd Epilog

Celebrating Excellence

Keynote by Daniela Borgmann
„Applying Energy Management and Predictive Maintenance using Edge and Cloud Computing "

33rd Epilog

Celebrating Excellence

- Distinguished Young Alumn Award
- Siemens Awards for Excellence
- Bachelor with Honors
- Best Poster Award

TU WIEN Informatics

Siemens Awards for Excellence

Together with Siemens Austria we honor the best female students of the year for outstanding achievements in their studies.

Siemens Awards for Excellence

Bachelor Students

- Lisa Fürst
- Sonja Gerstenmayer
- Sophia Schober
- Christina Tüchler

Master Students

- Jasmine Haider
- Theresa Müller
- Raffaela Pichler
- Timea Toth

33rd Epilog

Celebrating Excellence

Distinguished Young Alumn Award

Siemens Awards for Excellence

Bachelor with Honors

Best Poster Award

Bachelor with Honors

With our excellence program, we enable you to prove your potential in individually designed challenges, nurture your talents, and engage you in scientific research.

Bachelor with Honors

The Graduates

- **Andreas Auer**
  Software and Information Engineering
- **Adrian Chmurovic**
  Software and Information Engineering
- **Alexander Falzberger**
  Software and Information Engineering
- **Christoph Ullrich Günther**
  Software and Information Engineering

33rd Epilog

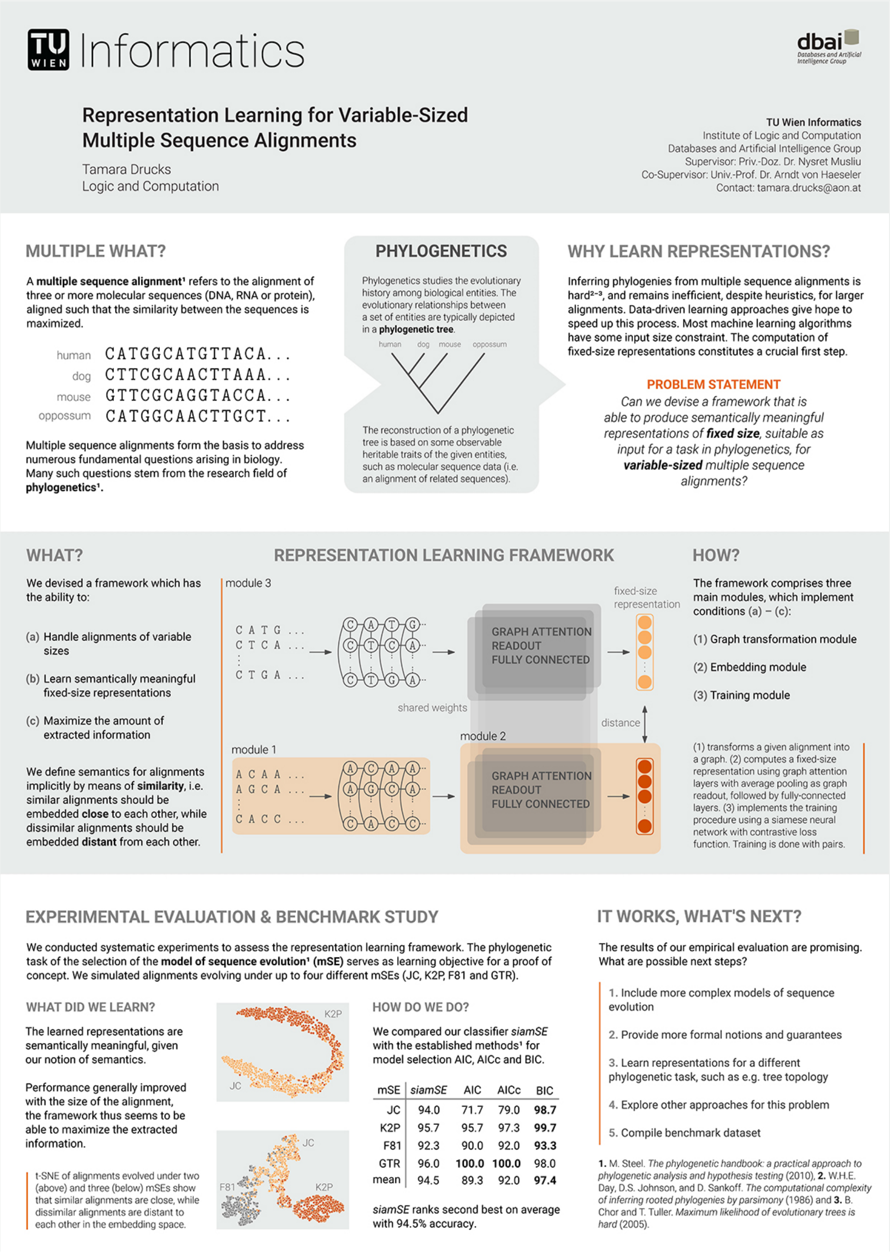
Celebrating Excellence

Distinguished Young Alumn Award

Siemens Awards for Excellence

Bachelor with Honors

Best Poster Award



Representation Learning for Variable-Sized Multiple Sequence Alignments

Tamara Drucks
Logic and Computation

TU Wien Informatics
Institute of Logic and Computation
Databases and Artificial Intelligence Group
Supervisor: Priv.-Doz. Dr. Nysret Musliu
Co-Supervisor: Univ.-Prof. Dr. Arndt von Haeseler
Contact: tamara.drucks@aon.at

MULTIPLE WHAT?

A **multiple sequence alignment**[1] refers to the alignment of three or more molecular sequences (DNA, RNA or protein), aligned such that the similarity between the sequences is maximized.

human	CATGGCATGTTACA...
dog	CTTCGCAACTTAAA...
mouse	GTTCGCAGGTACCA...
oppossum	CATGGCAACTTGCT...

Multiple sequence alignments form the basis to address numerous fundamental questions arising in biology. Many such questions stem from the research field of **phylogenetics**[1].

PHYLOGENETICS

Phylogenetics studies the evolutionary history among biological entities. The evolutionary relationships between a set of entities are typically depicted in a **phylogenetic tree**.

The reconstruction of a phylogenetic tree is based on some observable heritable traits of the given entities, such as molecular sequence data (i.e. an alignment of related sequences).

WHY LEARN REPRESENTATIONS?

Inferring phylogenies from multiple sequence alignments is hard[2-3], and remains inefficient, despite heuristics, for larger alignments. Data-driven learning approaches give hope to speed up this process. Most machine learning algorithms have some input size constraint. The computation of fixed-size representations constitutes a crucial first step.

PROBLEM STATEMENT

*Can we devise a framework that is able to produce semantically meaningful representations of **fixed size**, suitable as input for a task in phylogenetics, for **variable-sized** multiple sequence alignments?*

WHAT?

We devised a framework which has the ability to:

(a) Handle alignments of variable sizes

(b) Learn semantically meaningful fixed-size representations

(c) Maximize the amount of extracted information

We define semantics for alignments implicitly by means of **similarity**, i.e. similar alignments should be embedded **close** to each other, while dissimilar alignments should be embedded **distant** from each other.

REPRESENTATION LEARNING FRAMEWORK

HOW?

The framework comprises three main modules, which implement conditions (a) – (c):

(1) Graph transformation module

(2) Embedding module

(3) Training module

(1) transforms a given alignment into a graph. (2) computes a fixed-size representation using graph attention layers with average pooling as graph readout, followed by fully-connected layers. (3) implements the training procedure using a siamese neural network with contrastive loss function. Training is done with pairs.

EXPERIMENTAL EVALUATION & BENCHMARK STUDY

We conducted systematic experiments to assess the representation learning framework. The phylogenetic task of the selection of the **model of sequence evolution**[1] **(mSE)** serves as learning objective for a proof of concept. We simulated alignments evolving under up to four different mSEs (JC, K2P, F81 and GTR).

WHAT DID WE LEARN?

The learned representations are semantically meaningful, given our notion of semantics.

Performance generally improved with the size of the alignment, the framework thus seems to be able to maximize the extracted information.

t-SNE of alignments evolved under two (above) and three (below) mSEs show that similar alignments are close, while dissimilar alignments are distant to each other in the embedding space.

HOW DO WE DO?

We compared our classifier *siamSE* with the established methods[1] for model selection AIC, AICc and BIC.

mSE	*siamSE*	AIC	AICc	BIC
JC	94.0	71.7	79.0	**98.7**
K2P	95.7	95.7	97.3	**99.7**
F81	92.3	90.0	92.0	**93.3**
GTR	96.0	**100.0**	**100.0**	98.0
mean	94.5	89.3	92.0	**97.4**

siamSE ranks second best on average with 94.5% accuracy.

IT WORKS, WHAT'S NEXT?

The results of our empirical evaluation are promising. What are possible next steps?

1. Include more complex models of sequence evolution
2. Provide more formal notions and guarantees
3. Learn representations for a different phylogenetic task, such as e.g. tree topology
4. Explore other approaches for this problem
5. Compile benchmark dataset

1. M. Steel. *The phylogenetic handbook: a practical approach to phylogenetic analysis and hypothesis testing* (2010), **2.** W.H.E. Day, D.S. Johnson, and D. Sankoff. *The computational complexity of inferring rooted phylogenies by parsimony* (1986) and **3.** B. Chor and T. Tuller. *Maximum likelihood of evolutionary trees is hard* (2005).

Best Poster Award Winner

Tamara Drucks
Supervisor: Nysret Musliu
192-02 Research Unit of Databases and Artificial Intelligence

33rd Epilog

Celebrating Excellence

Distinguished Young Alumn Award

Siemens Awards for Excellence

Bachelor with Honors

Best Poster Award

TU WIEN Informatics

Distinguished Young Alumn Award

Candidates

- Axel Brunnbauer
- Tobias Geibinger
- Viktoriia Korchemna
- Sebastian Mazza

Distinguished Young Alumn Award Winner

Axel Brunnbauer

Model-Based Reinforcement Learning for Autonomous Racing

Supervisor: Radu Grosu
191-01 Cyber-Physical Systems

Distinguished Young Alumn Award Winner

Tobias Geibinger

Investigating Constraint Programming and Hybrid Answer-set Solving for Industrial Test Laboratory Scheduling

Supervisor: Nysret Musliu
192-02 Databases and Artificial Intelligence

Distinguished Young Alumn Award Winner

Viktoriia Korchemna

The Complexity of Bayesian Network Learning: Revisiting the Superstructure

Supervisor: Robert Ganian
192-01 Algorithms and Complexity

Distinguished Young Alumn Award Winner

Sebastian Mazza

Homomorphic-Encrypted Volume Rendering

Supervisor: Ivan Viola
193-02 Computer Graphics

33rd Epilog

Celebrating Excellence

Distinguished Young Alumn Award

Siemens Awards for Excellence

Bachelor with Honors

Best Poster Award

Informatics