

# MARTIN STEINEGGER

+1 · (443) · 627 · 0731 ◇ themartinsteinegger@gmail.com

502-423 1 Gwanak-ro, Gwanak-gu, ◇ Seoul, Korea

Born 29.01.1985 in Erding. German citizen.

## CURRICULUM VITAE

---

### Education

- 08/2014 - 08/2018 Ph.D. in Computer Science at the Technical University Munich (Passed with summa cum laude)
- 04/2013 - 08/2014 Master of Science in Computer Science at the Ludwig Maximilian University (Passed with merit)
- 09/2010 - 04/2013 Bachelor of Science in Bioinformatics at TU Munich / Ludwig Maximilian University
- 09/2006 - 07/2008 Business Informatics at the EDV-Schule Plattling (Passed as second best student)
- 10/2001 - 06/2005 Computer Engineering at the HTL Braunau (Technical college for electronics)

### Research and Industry experience

- since 03/2020 Assistant Professor at Seoul National University. *full-time*  
Bioinformatics lab
- 10/2018 - 02/2020 Postdoctoral Fellow at the Salzberg Lab at the Johns Hopkins University School of Medicine. *full-time*  
Pathogen detection in human metagenomic data.
- 08/2014 - 09/2018 PhD Student at the Quantitative and Computational Biology Laboratory at the Max-Planck Institute for Biophysical Chemistry. *full-time*  
Ultrafast and sensitive sequence search methods in the era of next generation sequencing.
- 04/2016 - 07/2018 Collaboration with Seok Lab at the Seoul National University.  
Large-scale de-novo structure prediction based on coevolution analysis of metagenomics-enriched multiple sequence alignments.
- 03/2015 - 04/2016 Collaboration with Notredame Lab at the Centre for Genomic Regulation in Barcelona.  
Supporting the development of a large scale multiple sequence aligner.
- 08/2014 - 12/2014 Visiting Scientist at the Seok Lab, Seoul National University. *full-time*  
Improving energy calculation for docking and protein structure prediction.
- 05/2012 - 07/2014 Research assistant at the Soeding Lab, Gene Center, LMU Munich. *part-time*  
Improving HMM remote homologues protein search method.
- 08/2013 - 10/2013 Visiting Scientist at the Sali Lab, UCSF. *full-time*  
Implementing a Bayesian inference framework to determine enzyme pathways.
- 07/2011 - 05/2012 Visiting Scientist at Rost Lab, Technical University Munich. *part-time*  
Full In-Silico mutagenesis of the human proteome using the Cloud.
- 06/2011 - 05/2012 Technical Architect / Scrum Master at Medability. *part-time*  
Developing a haptic surgery simulator
- 09/2008 - 06/2011 Software Engineer / Security Tester / Performance engineering at Accenture Technology Solutions. *full-time*
- 09/2007 - 01/2008 Software Engineer / Technical Architect at visionary people AG. *freelancer*
- 11/2005 - 06/2006 IT support at Bezirkskrankenhaus Haar. *full-time*

## ACHIEVEMENTS AND QUALIFICATIONS

---

### Awards, Fellowships and Achievements

2018	Poster award at the ECCB 2018
2016	Poster award at the Critical Assessment of Protein Structure Prediction 12 Conference
2015	Max Planck PhD fellowship
2013	Winner of the Twilio price (~1000\$) at the Disrupt TechCrunch Hackathon (1200+ attendees)
2012	Excellence initiative research grant, Ludwig Maximilian University
2012	AMD Research grant (~700\$ one graphic card)
2012	NVIDIA research grant (~3000\$ two graphic cards)
2011	Amazon research grant (10.000\$ Amazon Web Services credits)
2011	Finalist in the Big Data Challenge, CycleComputing
2008	Master prize of the Bavarian state government, EDV-Schule Plattling

### Certificates

2011	Certified ScrumMaster (CSM)
2010	ASDA Application Developer (Massachusetts Institute of Technology / Accenture)
2010	Information Technology Infrastructure Library V3 Foundation
2010	SpringSource Certified Spring Professional
2010	ISTQB Certified Tester
2009	Sun Certified Java Programmer
2008	IBM Certified System Administrator

### Technical Strengths

Programming	C++, C, Java, Shell scripting, Python
Cloud	Amazon Web Services
Testing	Unit, performance, functional, penetration test
Databases	BI, SQL, PL/SQL, Oracle 11g, MySQL, db4o

### Languages

German	Native
English	Fluent
Korean	Beginner

### Invited Peer Reviews

Science (2), Bioinformatics (5), PROTEINS (1)

### Public source code

Plass	Main developer. <a href="https://github.com/soedinglab/plass">github.com/soedinglab/plass</a>
Linclust	Main developer. <a href="https://github.com/soedinglab/mmseqs2">github.com/soedinglab/mmseqs2</a>
MMseqs2	Main developer. <a href="https://github.com/soedinglab/mmseqs2">github.com/soedinglab/mmseqs2</a>
MMseqs	Main developer with Maria Hauser. <a href="https://github.com/soedinglab/mmseqs">github.com/soedinglab/mmseqs</a>
HH-suite	Developing faster core algorithms. <a href="https://github.com/soedinglab/hh-suite">github.com/soedinglab/hh-suite</a>

## TALKS, POSTERS, AND PUBLICATIONS

---

### Talks

- 09/2019 University of Salzburg, Austria, New algorithms and tools for large-scale sequence analysis of metagenomic data
- 05/2019 University of Konstanz, Germany, New algorithms and tools for large-scale sequence analysis of metagenomic data
- 04/2019 RECOMB-SEQ 2019, USA, New algorithms and tools for large-scale sequence analysis of metagenomics data
- 01/2019 Seoul National University, Republic of Korea, Metagenomics data analysis on steroids
- 10/2018 Johns Hopkins University, USA, Metagenomics data analysis on steroids
- 09/2018 Max Planck Institute for Marine Microbiology, Germany, Metagenomics data analysis on steroids
- 07/2018 BiATA 2018, Russia, New algorithms and tools for large-scale sequence analysis of metagenomics data
- 07/2018 ISMB 2018, USA, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets
- 04/2018 European Bioinformatics Institute, England, Fast and sensitive protein sequence search, clustering and assembly tools for the analysis of massive metagenomics datasets
- 04/2018 NGS 2018, Spain, Fast and sensitive protein sequence search, clustering and assembly tools for the analysis of massive metagenomics datasets
- 01/2018 Johns Hopkins University, USA, Search, Clustering and Assembly tools for huge metagenomics datasets
- 01/2018 Rutgers University, USA, Search, Clustering and Assembly tools for huge metagenomics datasets
- 05/2017 Tokyo University, Japan, MMseqs2 / Linclust
- 05/2017 National Institute of Advanced Industrial Science, Japan, MMseqs2 / Linclust
- 06/2016 SocBIN2016, Russia, Sensitive protein sequence searching for the analysis of massive data sets
- 06/2015 Beijing Genomics Institute, China, HH-suite for sensitive protein sequence searching. / MMseqs for protein search
- 05/2015 Quest for Orthologs 4, Spain, MMseqs for clustering huge protein sets
- 03/2015 European Bioinformatics Institute, England, Sequence clustering and search in the era of NGS
- 06/2014 ISCB NGS14, Spain, MMseqs suite for fast and sensitive batch searching
- 06/2014 Hadoop User Group, Germany, In-Silico mutagenesis on Amazon EMR
- 09/2012 GMDS, Germany, Cloud architecture for In-Silico mutagenesis
- 12/2011 EDAM Meeting, Netherlands, Cloud architecture for PredictProtein
- 07/2010 University Cologne, Germany, Application Security
- 01/2010 Accenture community meeting, Germany, Web security

## Poster

- 11/2019 Genome Informatics 2019, USA, Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank
- 11/2019 Genome Informatics 2019, USA, New algorithms and tools for large-scale sequence analysis of metagenomic data
- 09/2018 ECCB18, USA, MMseqs2 desktop and local web server app for fast, interactive sequence searches
- 07/2018 ISMB 2018, USA, MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets
- 04/2017 ISMB NGS 2017, Spain, Sensitive protein sequence searching for the analysis of massive data sets
- 12/2016 CASP12, Italy, Sensitive protein sequence searching for the analysis of massive data sets
- 04/2016 ISMB NGS 2016, Spain, Sensitive protein sequence searching for the analysis of massive data sets
- 03/2016 ABLIS 2016, Belgium, Fast and sensitive searching of proteomic data
- 05/2015 Quest for Orthologs 4, Spain, MMseqs for clustering huge protein sets
- 09/2014 KIAS Conference on Protein Structure and Function, Republic of Korea, Accelerated pairwise HMM alignment using SIMD programming and improved secondary structure scoring

## Publications

The most important articles are highlighted in red.

- [1] **Steinegger, M.**, Salzberg L S. (2020) Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *bioRxiv*, doi: 10.1101/2020.01.26.920173

## Peer-reviewed manuscripts

- [1] **Steinegger, M.**, Markus Meier, Milot Mirdita, Harald Vöhringer, Stephan J. Haunsberger, and Söding, J. (2019) HH-suite3 for fast remote homology detection and deep protein annotation. *BMC Bioinformatics*, doi: 10.1186/s12859-019-3019-7
- [2] **Steinegger, M.**, Milot Mirdita, and Söding, J. (2019) Protein-level assembly increases protein sequence recovery from metagenomic samples manifold *Nature Methods*, **16**, 603–606, doi: 10.1038/s41592-019-0437-4
- [3] Milot Mirdita, **Steinegger, M.** and Söding, J. (2019) MMseqs2 desktop and local web server app for fast, interactive sequence searches. *Bioinformatics*. doi:10.1093/bioinformatics/bty1057
- [4] **Steinegger, M.**, and Söding, J. (2018) Clustering huge protein sequence sets in linear time *Nature Communications* doi: 10.1038/s41467-018-04964-5
- [5] Mahlich Y., **Steinegger, M.**, Rost, B. and Bromberg Y. (2018) HFSP: High speed homology-driven function annotation of proteins. *Bioinformatics*, bf 34, i304–i312, doi: 10.1093/bioinformatics/bty262
- [6] **Steinegger, M.**, and Söding, J. (2017) MMseqs2: Sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, **35**, 1026–1028, doi: 10.1038/nbt.3988
- [7] Mirdita, M.<sup>#</sup>, von den Driesch<sup>#</sup>, L., Galiez, G., Martin, M., Söding, J.<sup>\*</sup>, and **Steinegger, M.**<sup>\*</sup> (2017) Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Research*, **45**, D170–D176, doi: 10.1093/nar/gkw1081. . (<sup>#</sup>Equal contributions.) (<sup>\*</sup>Corresponding authors.)
- [8] Hauser M.<sup>#</sup>, **Steinegger, M.**<sup>#</sup>, and Söding, J. (2016) MMseqs software suite for fast and deep clustering and searching of large protein sequence sets *Bioinformatics*, **32**, 1323-1330. doi: 10.1093/bioinformatics/btw006. (<sup>#</sup>Equal contributions.)

- [9] Kajan L., Yachdav G., Vicedo E., **Steinegger M.**, Mirdita M., Angermüller C., Böhm A., Domke S., Ertl J., Mertens C., Reisinger E., Staniewski C., B. Rost (2014) Cloud prediction of protein structure and function with PredictProtein for Debian *BioMed research international*, doi: 10.1155/2013/398968

**Non peer-reviewed articles**

- [1] **Steinegger M.** and Goiss, H. (2011) Introducing a Model-based Automated Test Script Generator *Testing Experience*, 70-76

## TEACHING

---

### Lectures, seminars, and lab classes

- 2018 Modern and scalable tools for efficient analysis of very large metagenomic at ECCB18 in Greece.
- 2012 Bioinformatics tutorial for bachelor students: Development of tutorial material and teaching at the Ludwig Maximilian University.
- 2009 - 2011 Database faculty at Accenture. Regularly held Oracle database seminars and reworked the course material. Full-time 2 day seminars for Accenture consultants
- 2010 - 2011 Security training at Accenture. Helped create a security curriculum and held seminars.
- 2010 Java architecture seminars at Accenture, Full-time 5 days workshop for Java consultants

### Co-Supervised theses

03/2016 - 09/2016: Milot Mirdita, Computer Science, LMU Munich  
*Uniclust - clustered and deeply annotated protein sequence databases*

04/2014 - 10/2014: Lars von der Driesch, Bioinformatics, LMU Munich / TU Munich  
*Deep clustering and annotation of the Uniprot database*

11/2013 - 05/2014: Stefan Haunsberger, Bioinformatics, Hochschule Weihenstephan-Triesdorf  
*Fast AVX-based Forward-Backward and Maximum Accuracy algorithms for pairwise alignment of profile hidden Markov models*