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* Laboratory of Machine Learning & Bioinformatics
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	NVIDA 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

Public source code

ColabFold	https://github.com/sokrypton/ColabFold
Foldseek	https://github.com/steineggerlab/foldseek
Conterminator	https://github.com/martin-steinegger/conterminator
Plass	github.com/soedinglab/plass
Linclust	github.com/soedinglab/mmseqs2
MMseqs2	github.com/soedinglab/mmseqs2
MMseqs	github.com/soedinglab/mmseqs
HH-suite	github.com/soedinglab/hh-suite

TALKS, POSTERS, AND PUBLICATIONS

Talks

11/2021	Swiss Institute of Bioinformatics, Switzerland, Next generation protein analysis tools in the ear of highly accurate protein structure prediction
11/2021	KSMCB 2021, Korea, Mega scale protein structure prediction and search
08/2021	Boston Protein Design and Modeling Club, USA, ColabFold - Making protein folding accessible to all via Google Colab!
07/2021	BiATA Conference, Russia, MMseqs2 profile/profile: fast and ultra sensitive searches beyond the twilight zone
06/2021	BVCN Conference, USA, Metagenomic pathogen detection using MMseqs2, Plass, and Linclust
12/2020	MicroEvo Meeting Informatics, Denmark, The unresolved dying of the Mariana crows
09/2020	Genome Informatics, UK, Protein-guided nucleotide viral genome assembly for huge metagenomic datasets
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 ear of NGS
06/2014	ISCB NGS14, Spain, MMseqs suite for fast and sensitive batch searching
06/2014	Hadoop User Group, Germany, In-Silico mutagenises 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	ABLS 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 programing and improved secondary structure scoring

Publications

The most important articles are highlighted in red.

[1] Liu D., and **Steinegger M.** (2021), Block aligner: fast and flexible pairwise sequence alignment with SIMD-accelerated adaptive blocks *bioRxiv*, doi: 10.1101/2021.08.15.456425

[2] Mirdita M., Schütze K., Moriwaki Y., Heo L.,Ovchinnikov S. and **Steinegger M.** (2021), ColabFold - Making protein folding accessible to all *bioRxiv*, doi: 10.1101/2021.08.15.456425

[3] Vanni, C., Schechter, M., Delmont, T., and others (2021), AGNOSTOS-DB: a resource to unlock the uncharted regions of the coding sequence space *bioRxiv*, doi: 10.1101/2021.06.07.447314

[4] Vanni, C., Schechter, M., Silvia G., Barberán, A, Buttigieg, P., Casamayor, E., Delmont, T., Duarte, C., Eren, A. and Finn, R. and others (2021), Unifying the global coding sequence space enables the study of genes with unknown function across biomes *bioRxiv*, doi: 10.1101/2020.06.30.180448

Peer-reviewed manuscripts

[1] Pockrandt, C., **Steinegger M.**, Salzberg L S. (2021), PhyloCSF++: A fast and user-friendly implementation of PhyloCSF with annotation tools. *Bioinformatics*, doi: 10.1093/bioinformatics/btab756

[2] Jumper, J., Evans, R., Pritzel, A., Green, T. and others (2021), Applying and improving AlphaFold at CASP14 *Proteins: Structure, Function, and Bioinformatics*, doi: 10.1002/prot.26257

[3] Jumper J., Evans R., Pritzel A., Green T. and others (2021), Highly accurate protein structure prediction with AlphaFold. *Nature*, doi: 10.1038/s41586-021-03819-2

[4] Elnaggar, A., Heinzinger, M., Dallago, C., Rehawi, G., Wang, Y., Jones, L., Gibbs, T., Feher, T., Angerer, C., **Steinegger, M.** and others (2021), ProtTrans: Towards Cracking the Language of Lifes Code Through Self-Supervised Deep Learning and High Performance Computing, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, doi: 10.1109/TPAMI.2021.3095381

[5] Aevarsson, A., Kaczorowska, A., Adalsteinsson, A., Ahlqvist, J., Al-Karadaghi, S., and others (2021), Going to extremes – a metagenomic journey into the dark matter of life *FEMS Microbiology Letters* 10.1093/femsle/fnab067

[6] Michael Bernhofer, Christian Dallago, Tim Karl, Venkata Satagopam, Michael Heinzinger, Maria Littmann, Tobias Olenyi, Jiajun Qiu, Konstantin Schuetze, Guy Yachdav, Haim Ashkenazy, Nir Ben-Tal, Yana Bromberg, and others (2021), PredictProtein-Predicting Protein Structure and Function for 29 Years, *Nucleic Acids Research*, doi: 10.1093/nar/gkab354

[7] Mirdita, M. and **Steingger, M.**, and Breitwieser, F. and Soeding, J. and Karin, E. L. (2021), Fast and sensitive taxonomic assignment to metagenomic contigs. *Bioinformatics*, doi: 10.1101/2020.11.27.401018

[8] Credle, J. J., Robinson, M., Gunn, J., Monaco, D., Sie, B., Tchir, A. L., Hardick, J., Zheng, X., Shaw-Saliba, K., and Rothman, Richard and others (2021), Highly multiplexed oligonucleotide probe-ligation testing enables efficient extraction-free SARS-CoV-2 detection and viral genotyping. *Modern Pathology*, doi: 10.1038/s41379-020-00730-5

[9] Zhao, Bi and Katuwawala, Akila and Oldfield, Christopher J and Dunker, A Keith and Faraggi, Eshel and Gsponer, Jörg and Kloczkowski, Andrzej and Malhis, Nawar and Mirdita, Milot and Obradovic, Zoran and others (2021), DescribePROT: database of amino acid-level protein structure and function predictions. *Nucleic Acids Research*, doi: 10.1093/nar/gkaa931

[10] Gabler F., Nam S., Till S., Mirdita M., **Steinegger M.**, Söding J., Lupas A, Alva V., (2020), Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. *Current Protocols in Bioinformatics*, doi: 10.1002/cpbi.108

[11] Park S., **Steinegger, M.**, Cho H. and Chun J. (2020) Metagenomic Association Analysis of Gut Symbiont Limosilactobacillus reuteri Without Host-Specific Genome Isolation. *Frontiers in Microbiology*, doi: 10.3389/fmicb.2020.585622

[12] **Steinegger, M.**, Salzberg L S. (2020) Terminating contamination: large-scale search identifies more than 2,000,000 contaminated entries in GenBank. *Genome Biology* , doi: 10.1186/s13059-020-02023-1

[13] **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

[14] **Steinegger, M.**, Milot Mirdita, and Söding, J. (2019) Protein-level assembly increases protein sequence recovery from metagenomic samples manyfold. *Nature Methods*, **16**, 603–606, doi: 10.1038/s41592-019-0437-4

[15] 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

[16] **Steinegger, M.**, and Söding, J. (2018) Clustering huge protein sequence sets in linear time. *Nature Communications* doi: 10.1038/s41467-018-04964-5

[17] Forslund K., Pereira C., Capella-Gutierrez S. and others (2018) Gearing up to handle the mosaic nature of life in the quest for orthologs *Bioinformatics*, bf 34, i323–i329, doi: 10.1093/bioinformatics/btx542

[18] 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

[19] **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

[20] Mirdita, M.[#], von den Driesch[#], L., Galiez, G., Martin, M., Söding, J.[*], and **Steinegger, M.**[*] (2017) Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Research*, **45**, D170–D176, doi: 10.1093/nar/gkw1081. . ([#]Equal contributions.) ([*]Corresponding authors.)

[21] Hauser M.[#], **Steinegger, M.**[#], 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. ([#]Equal contributions.)

[22] Kajan L., Yachdav G., Vicedo E., **Steinegger M.**, Mirdita M., Angermüller C., Böhm A., Domke S., Ertl J., Mertes 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

2020	Deep dive into metagenomic data using metagenome-atlas and MMseqs2 at ECCB 2020 in Spain.
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