

Dave Messina.

4100 Forest Park #217
Saint Louis, MO 63108
314.974.6377

<http://davemessina.com>
dave@davemessina.com

Education

2012 **PhD, Computational Biology**, Stockholm University
2004 **MA, Genetics**, Washington University School of Medicine
1996 **BA Biology**, University of Illinois at Chicago

Research Experience

Director of Analysis @ Cofactor Genomics (2012-present)

- Leads Cofactor's computational research & development program, involving software and infrastructure
- Oversees bioinformatics analyses for all clients, from small academic labs to large pharma companies
- Represents the company in both pre- and post-sales customer contacts

PhD candidate @ Stockholm Bioinformatics Centre, Stockholm University (2007-2012)

- Thesis project: developing methods to identify novel protein-coding genes from metagenomic samples
- Created DASHer, a Java viewer for integrating distributed protein annotations (dasher.sbc.su.se)
- Co-designed and developed the API for SeqXML and OrthoXML, formats for comparative orthology (seqxml.org)

Core Developer @ BioPerl (2006-present)

- Co-wrote the Deobfuscator, a highly-used documentation tool (bioperl.org/wiki/Deobfuscator)
- Maintain, debug, and improve the codebase for a large open-source project (~700,000 lines of code)
- Provide frontline support to user community
- Plan and guide the future direction of the project

Senior Programmer Analyst @ Genome Sequencing Center, Washington University (2004-2007)

- Created semblance, an open-source program for fully-automated quality-control analysis and comparison of whole-genome-shotgun assemblies
- Designed and implemented an EST-cluster viewer for the Parasitic Nematode Project (www.nematode.net)
- Assessed next-generation sequencing platforms and developed analysis pipeline prototypes

Graduate student @ Department of Genetics, Washington University (1998-2003)

- Designed and built a microarray of the (nearly) complete set of human transcription factor genes
- Employed expression profiling analysis methods for the first large-scale expression study of human craniofacial development

Computational Experience

- 10 years of experience in algorithm design, sequence analysis, expression analysis, comparative genomics, next-gen sequencing, and visualization tools
- 7 years of experience designing and building large-scale systems for high-throughput analysis
- 14 years of programming: Perl, Java, and a little C. Experience in implementing parallel processing
- Trained in sequence alignment algorithms, gene prediction, structure prediction, and database design
- Assistant Instructor for the Cold Spring Harbor Programming for Biologists course for the last 5 years (meetings.cshl.edu/courses/c-info11.shtml)

Publications

Prediction of novel protein families from metagenomic samples.

Messina DN, Lysholm F, Andersson B, and Sonnhammer ELL. 2012. *submitted*.

SeqXML and OrthoXML: standards for sequence and orthology information.

Schmitt T, **Messina DN**, Schreiber F, Sonnhammer ELL. 2011. *Briefings in Bioinformatics*.

<http://dx.doi.org/10.1093/bib/bbr025>

InParanoid 7: new algorithms and tools for eukaryotic orthology analysis.

Östlund G, Schmitt T, Forslund K, Köstler T, **Messina DN**, Roopra S, Frings O, Sonnhammer ELL. 2010.

Nucleic Acids Res 38:D196-203. <http://dx.doi.org/10.1093/nar/gkp931>

MetaTM - a consensus method for transmembrane protein topology prediction.

Klammer M, **Messina DN**, Schmitt T, Sonnhammer ELL. 2009. *BMC Bioinformatics* 10:314.

<http://dx.doi.org/10.1186/1471-2105-10-314>

DASher: a stand-alone protein sequence client for DAS, the Distributed Annotation System.

Messina DN and Sonnhammer ELL. 2009. *Bioinformatics* 25:1333.

<http://dx.doi.org/10.1093/bioinformatics/btp153>

NemaPath: online exploration of KEGG-based metabolic pathways for nematodes.

Wylie T, Martin J, Abubucker S, Yin Y, **Messina DN**, Wang Z, McCarter JP, Mitreva M. 2008. *BMC*

Genomics 9:525. <http://dx.doi.org/10.1186/1471-2164-9-525>

Evolutionary and biomedical insights from the rhesus macaque genome.

Rhesus Macaque Genome Sequencing and Analysis Consortium. 2007. *Science* 316:222-234.

<http://dx.doi.org/10.1126/science.1139247>

Physical map-assisted whole-genome shotgun sequence assemblies.

Warren RL, Varabei D, Platt D, Huang X, **Messina D**, Yang S-P, Kronstad J, Krzywinski M, Warren W,

Wallis J, Hillier L, Chinwalla A, Schein J, Siddiqui A, Marra M, Wilson R, Jones S. 2006. *Genome Res*

16:768-775. <http://dx.doi.org/10.1101/gr.5090606>

Gene expression in pharyngeal arch 1 during human embryonic development.

Cai J, Ash D, Kotch LE, Jabs EW, Attie-Bitach T, Auge J, Mattei G, Etchevers H, Vekemans M,

Korshunova Y, Tidwell R, **Messina DN**, Winston JB, and Lovett M. 2005. *Hum Mol Genet* Apr 1;14(7):

903-12. Epub 2005 Feb 9. <http://dx.doi.org/10.1093/hmg/ddi083>

An ORFeome-Based Analysis of Human Transcription Factor Genes and the Construction of a Microarray to Interrogate their Expression.

Messina DN, Glasscock J, Gish W, and Lovett M. 2004. *Genome Res* Oct;14(10B):2041-7.

<http://dx.doi.org/10.1101/gr.2584104>

Recipes for Success? A Review of Bioinformatics:Methods and Protocols.

Messina DN. 2000. *Trends in Biotechnol* 18:224-225.

[http://dx.doi.org/10.1016/s0167-7799\(00\)01435-9](http://dx.doi.org/10.1016/s0167-7799(00)01435-9)

Linkage of familial dilated cardiomyopathy with conduction defect and muscular dystrophy to chromosome 6q23.

Messina DN, Speer MC, Pericak-Vance MA, and McNally EM. 1997. *Am J Hum Genet* 61:909-917.

<http://dx.doi.org/10.1086%2F514896>