

Frederick Albert Matsen IV

Computational Biology Program
Fred Hutchinson Cancer Research Center
1100 Fairview Ave. N., M1-B855
P.O. Box 19024
Seattle, WA 98109-1024
(206) 667-7318
matsen@fredhutch.org
<http://matsen.fredhutch.org/>

530 Broadway E #230
Seattle, WA 98102

- Positions**
- Associate Member**, Fred Hutchinson Cancer Research Center, Seattle, Washington.
April 2015 – Present
 - Affiliate Associate Professor**, Statistics, University of Washington, Seattle, Washington.
June 2016 – Present
 - Assistant Member**, Fred Hutchinson Cancer Research Center, Seattle, Washington.
August 2010 – March 2015
 - Miller Postdoctoral Scholar**, UC Berkeley, Berkeley, California.
July 2007 – July 2010
 - Allan Wilson Postdoctoral Scholar**, University of Canterbury, Christchurch, New Zealand.
July 2006 – June 2007
- Education**
- Harvard University**, Cambridge, MA.
Ph.D. in Mathematics, June 2006.
 - Université de Paris 13**, Paris, France.
D.E.A. in Mathematics, June 2001.
 - Stanford University**, Stanford, CA.
B.S. in Mathematics with honors, June 2000.
- Grants:**
- primary investigator**
 - HHMI-Simons Faculty Scholar Award** (11/01/2016 – 11/01/2021)
\$900,000 total cost
 - NSF Award 1564137** (07/01/2016 – 06/30/2020)
Enabling phylogenetic inference for modern data sets.
Co-PI with Vladimir Minin. \$824,898 total cost
 - R01 GM113246-01** (08/01/2014 – 08/01/2019)
Leveraging deep sequencing data to understand antibody maturation.
\$1,888,860 total cost
 - NSF Award 1223057** (09/01/2012 – 08/31/2017)
New theorems and algorithms for comprehensive analysis of metagenomic data via statistical phylogenetics.
Co-PI with Aaron Darling. \$812,615 total cost
 - R01 HG005966** (09/27/2010 – 06/30/2013)
Novel Computational Tools for Studying the Human Microbiome.
Co-PI with David Fredricks. \$1,520,743 total cost
 - Support for <http://phyloseminar.org/>** (08/22/2013 – present)
From the Society of Systematic Biologists.

ITHS pilot grant (02/01/14 – 02/01/15)

Inferring statistical models of antibody somatic hypermutation from high-throughput sequencing data

CFAR New Investigator Award (01/01/14 – 01/01/16)

Machine learning, informed by structural knowledge, to expand our knowledge of APOBEC3 specificity and improve hypermutation inference

Grant support: **R01-AI120961** (2016/01/01 – 2020/12/31)

co-investigator Defining the Infant Immune Response to HIV.

PI: Julie Overbaugh.

Our role: our group will design and perform all the B cell receptor sequence analysis for this project.

Bill and Melinda Gates Foundation OPP1110049 (08/31/2014 – 7/31/2016)

Biostatistical, Computational Biology, and Mathematical Modeling for the Assessment of Immune Correlates of Protection in the HVTN 701 and 702 Efficacy Trials in South Africa.

PI: Peter Gilbert.

Our role: Improve methods for reconstruction of HIV molecular sequences which found infection using later time point samples.

R37 AI038518-16A2 (03/01/2010 – 02/28/2015)

Early and Reinfection in High Risk Women.

PI: Julie Overbaugh.

Our role: our group has designed and performed all the bioinformatics for this project.

R21 AI104449-01 (02/15/13 – 01/31/15)

PrEP Exposure during HIV-1 Infection and the Risk of Low-Frequency ARV Resistance.

PI: Dara Lehman.

Our role: our group has designed and performed all the bioinformatics for this project.

U54GM111274-01 (09/12/2014 – 06/30/2019)

Center for Statistics and Quantitative Infectious Diseases.

PI: Elizabeth Halloran.

Our role: design and implement computational phylogenetic algorithms for pathogen sequence data.

Awards

HHMI-Simons faculty scholar (November 2016 – October 2021)

Research support for five years.

Miller fellow (July 2007 – July 2010)

Support for three years of postdoctoral research at UC Berkeley.

Allan Wilson postdoctoral fellow (July 2006 – June 2007)

Support for one year of postdoctoral research.

NSF graduate research fellow (July 2003 – June 2006)

Support for three years of study.

Harvard University certificate of distinction in teaching (Fall 2004)

For teaching second semester calculus.

Jean D. Valpine fellowship (September 2001 – June 2003)

Support for two years of study.

Fulbright fellowship (July 2000 – June 2001)

To study mathematics in Paris.

Hofer prize for excellence in undergraduate writing (July 1998)

For a mathematics paper in real analysis.

Research interests

phylogenetics, innate and adaptive immunity, infectious disease, human microbiome, Markov processes, coalescent theory, molecular evolution, macroevolutionary models, discrete geometry, combinatorics, Bayesian statistics, bioinformatics.

Frederick Albert Matsen IV

- Activities
- Invited session organizer: *Bernoulli Society World Congress in Probability and Statistics*, July 2016.
 - Invited session organizer: *Joint Evolution Meetings*, June 2016.
 - Mentor: Summer Undergraduate Research Program, Summer 2016.
 - Mentor: Summer High School Internship Program, Summer 2014-6.
 - Organizer: Spotlight session on next generation phylogenetic inference, Evolution meeting, 2016.
 - Associate editor: *PLOS Computational Biology*, since December 2015.
 - Editor: Special issue of *Philosophical Transactions of the Royal Society B* on antibody dynamics, 2015.
 - Invited program committee member: *International Conference on Intelligent Systems for Molecular Biology (ISMB) / European Conference on Computational Biology (ECCB)*, 2009 and 2015.
 - Member: Statistical Genetics faculty group at the University of Washington, since October 2014.
 - Founder: *fredhutch.io* initiative to improve access to and education on scientific computing at the Fred Hutch (<http://fredhutch.io/>), since July 2014.
 - Founder and organizer: *Phylobabble* online phylogenetics discussion forum (<http://phylobabble.org/>), since March 2014.
 - Associate editor: *BMC Bioinformatics*, since July 2013.
 - Editorial board: *Systematic Biology*, since February 2013.
 - Host for professor on sabbatical: Vladimir Minin, academic year 2013-4.
 - Founder and organizer: *Phyloseminar* online phylogenetics seminar (<http://phyloseminar.org/>), since November 2009.
 - Contributor: *Biopython project* (<http://github.com/biopython/>), since December 2010.
 - Conference co-organizer: *Molecular Evolution and Phylogenetics*, Statistical and Applied Mathematical Sciences Institute (SAMSI), Spring 2009.
 - Paper and grant reviewer (selected list): *Annals of Human Genetics*, *Bioinformatics*, *Journal of Bioinformatics and Computational Biology*, *Bulletin of Mathematical Biology*, *ISMB/EECB 2015*, *Journal of Mathematical Biology*, *Journal of Molecular Evolution*, *Molecular Biology and Evolution*, *National Science Foundation*, *PLOS Pathogens*, *PNAS*, *STOC08*, *Systematic Biology*, *IEEE Transactions on Computational Biology and Bioinformatics*.
- Invited and conference presentations
- "Learning about antibody affinity maturation from sequence data using probabilistic models", *Penn Symposium on Mathematical & Computational Biology*, May 2017.
 - "Learning from immune repertoire sequences with probabilistic methods", *Stanford Computational and Systems Immunology Series*, August 2016.
 - "Consistency and convergence rate of phylogenetic inference via regularization", *Bernoulli Society World Congress in Probability and Statistics*, July 2016.
 - "The adaptive immune system: a grand and beautiful stochastic process", *University of Washington Statistics Seminar*, April 2016.
 - "Learning how antibodies are drafted and revised", *Simon Fraser University*, March 2016, *University of Southern California*, November 2015, *University of Pennsylvania*, November 2015, *National Bioterrorism and Countermeasures Center*, Frederick, MD, June 2015.
 - "Degree, curvature, and mixing of random walks on the phylogenetic subtree-prune-regraft graph", *Annual meeting of the Cladistics Society*, June 2015.
 - "Statistical inference of annotations and rearrangement groups and evolutionary constraints on B cell receptors", *Keystone Symposia on The Golden Anniversary of B Cell Discovery*, March 2015.

"Learning how antibodies are drafted and revised", *Center for Computational Biology Seminar Series*, UC Berkeley, April 2015.

"Leveraging the discrete and continuous structure of phylogenetic trees for the analysis of metagenomic data", *Joint Statistical Meetings*, August 2014.

"From the Ramayana to the Reverend Bayes: host defenses and zoonotic transmission of simian foamy virus", *Collège de France*, Paris, France, June 2013.

"Phylogenetics and the human microbiome", *LIRMM Molecular and Computational Biology Workshop*, Montpellier, France, June 2013.

"Analysis of metagenomes using phylogenetics" *American Society of Microbiology General Meeting*, Denver, CO, May 2013.

"Phylogenetics for environmental genomics" *Phycological Society of America Annual Meeting*, Seattle, WA, July 2011.

"Analysis and comparison of microbial sequence data via phylogenetic placements" *International Human Microbiome Congress*, Vancouver, BC, March 2011.

"Analysis and comparison of microbial sequence data via phylogenetic placements" *High-Throughput Biodiversity Research Using Eukaryotic Metagenetics*, NESCENT (Durham, NC), January 2011.

"The phylogenetic Kantorovich-Rubinstein metric for environmental DNA sequence samples" *Joint Statistical Meetings*, Vancouver, August 2010.

"From phytoplankta to rogue taxa: phylogenetics theory and practice" *Workshop in Biostatistics*, Stanford, February 2010.

"The phylogenetic geometry of rogue taxa" *New Mathematical Challenges from Molecular Biology and Genetics*, Banff International Research Station, Fall 2009.

"The phylogenetic geometry of rogue taxa" *Cyberinfrastructure for Phylogenetic Research (CIPRES) All Hands Meeting*, Berkeley, Fall 2009.

"Glimpses from the strange world of phylogenetic mixtures" *Future Directions in Phylogenetics*, Isaac Newton Institute for Mathematical Sciences, Fall 2008.

"Discrete harmonic analysis and the geometry of model-based phylogenetics" *Workshop on Interactions between Probability Theory and Computer Science*, Cornell, 2008.

Publications

Vu Dinh, Arman Bilge, Cheng Zhang, and Frederick A Matsen IV. Probabilistic Path Hamiltonian Monte Carlo. *arXiv*, 25 February 2017. <http://arxiv.org/abs/1702.07814>.

Alex Gavryushkin, Chris Whidden, and Frederick Matsen IV. The combinatorics of discrete time-trees: theory and open problems. <http://biorxiv.org/content/early/2016/10/27/063362>, 27 October 2016.

F Matsen, S Billey, A Kas, and M Konvalinka. Tanglegrams: a reduction tool for mathematical phylogenetics. *IEEE/ACM Trans. Comput. Biol. Bioinform.*, PP(99):1–1, 2016.

Vu Dinh, Aaron E Darling, and Frederick A Matsen IV. Online Bayesian phylogenetic inference: theoretical foundations via Sequential Monte Carlo. *arXiv; submitted to Systematic Biology*, 26 October 2016. <http://arxiv.org/abs/1610.08148>.

Chris Whidden and Frederick A Matsen IV. Efficiently inferring pairwise subtree Prune-and-Regraft adjacencies between phylogenetic trees. *arXiv*, 28 June 2016. <http://arxiv.org/abs/1606.08893>.

Vu Dinh, Lam Si Tung Ho, Marc A Suchard, and Frederick A Matsen IV. Consistency and convergence rate of phylogenetic inference via regularization. *arXiv; accepted to Annals of Statistics*, 9 June 2016. <http://arxiv.org/abs/1606.03059>.

Chris Whidden and Frederick A Matsen IV. Calculating the unrooted subtree prune-and-regraft distance. *arXiv*, 24 November 2015. <http://arxiv.org/abs/1511.07529>.

- Vu Dinh and Frederick A Matsen IV. The shape of the one-dimensional phylogenetic likelihood function. *In press, Annals of Applied Probability*, 13 July 2015. <http://arxiv.org/abs/1507.03647>.
- Duncan K Ralph and Frederick A Matsen IV. Likelihood-Based inference of B cell clonal families. *PLoS Comput. Biol.*, 12(10):e1005086, 17 October 2016.
- Sara C Billey, Matjaž Konvalinka, and Frederick A Matsen IV. On the enumeration of tanglegrams and tangled chains. *J. Combin. Theory Ser. A*, 146:239–263, 2017.
- Chris Whidden and Frederick A Matsen IV. Ricci-Ollivier curvature of the rooted phylogenetic Subtree-Prune-Regraft graph. In James Allen Fill and Mark Daniel Ward, editors, *2016 Proceedings of the Thirteenth Workshop on Analytic Algorithmics and Combinatorics (ANALCO)*, pages 106–120. SIAM, 2016.
- Duncan K Ralph and Frederick A Matsen IV. Consistency of VDJ rearrangement and substitution parameters enables accurate B cell receptor sequence annotation. *PLoS Comput. Biol.*, 12(1):e1004409, January 2016.
- Sarah Cobey, Patrick Wilson, and Frederick A Matsen IV. The evolution within us. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 370(1676), 5 September 2015.
- Connor O McCoy, Trevor Bedford, Vladimir N Minin, Philip Bradley, Harlan Robins, and Frederick A Matsen IV. Quantifying evolutionary constraints on B-cell affinity maturation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, 370(1676), 5 September 2015.
- Chris Whidden and Frederick A Matsen IV. Quantifying MCMC exploration of phylogenetic tree space. *Systematic Biology*, 64(3):472–491, May 2015.
- Alevtina Gall, Jutta Fero, Connor McCoy, Brian C Claywell, Carissa A Sanchez, Patricia L Blount, Xiaohong Li, Thomas L Vaughan, Frederick A Matsen, Brian J Reid, and Nina R Salama. Bacterial composition of the human upper gastrointestinal tract microbiome is dynamic and associated with genomic instability in a Barrett’s esophagus cohort. *PLoS One*, 10(6):e0129055, 15 June 2015.
- Dara A Lehman, Jared M Baeten, Connor O McCoy, Julie F Weis, Dylan Peterson, Gerald Mbarah, Deborah Donnell, Katherine K Thomas, Craig W Hendrix, Mark A Marzinke, Lisa Frenkel, Patrick Ndase, Nelly R Mugo, Connie Celum, Julie Overbaugh, Frederick A Matsen, and The Partners PrEP study team. Risk of drug resistance among persons acquiring HIV within a randomized clinical trial of single- or Dual-Agent preexposure prophylaxis. *J. Infect. Dis.*, 13 January 2015.
- Frederick A Matsen IV. Phylogenetics and the human microbiome. *Syst. Biol. (invited review)*, 64(1):e26–41, January 2015.
- Noah G Hoffman and Frederick A Matsen. Computational tools for taxonomic assignment. In *Encyclopedia of Metagenomics*, pages 1–6. Springer New York, 3 April 2014.
- Aaron E Darling, Guillaume Jospin, Eric Lowe, Frederick A Matsen IV, Holly M Bik, and Jonathan A Eisen. PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*, 2:e243, 9 January 2014. Selected as one of the top 20 articles from 2014 in PeerJ (<http://peerj.com/collections/5-peerjpicks/>).
- Frederick A Matsen IV, Christopher T Small, Khanh Soliven, Gregory A Engel, Mostafa M Feeroz, Xiaoxing Wang, Karen L Craig, M Kamrul Hasan, Michael Emerman, Maxine L Linial, and Lisa Jones-Engel. A novel Bayesian method for detection of APOBEC3-Mediated hypermutation and its application to zoonotic transmission of simian foamy viruses. *PLoS Comput. Biol.*, 10(2):e1003493, 27 February 2014.
- Khanh Soliven, Xiaoxing Wang, Christopher T Small, Mostafa M Feeroz, Eun-Gyung Lee, Karen L Craig, Kamrul Hasan, Gregory A Engel, Lisa Jones-Engel, Frederick A Matsen, and Maxine L. Linial. Simian foamy virus infection of rhesus macaques in Bangladesh: Relationship of latent proviruses and transcriptionally active viruses. *Journal of virology*, 87(24):13628–13639, 2013.

- L. Jones-Engel, G. Engel, C. Small, K. Soliven, M. Feeroz, X. Wang, M. Hasan, G. Oh, S. M. R. Alam, K. Craig, D. Jackson, F. Matsen IV, and M. L. Linial. Zoonotic simian foamy virus in Bangladesh reflects diverse patterns of transmission and co-infection. *Emerging Microbes and Infections*, 2(9):e58, 2013.
- K. Ronen, C. O. McCoy, F. A. Matsen, D. F. Boyd, S. Emery, K. Odem-Davis, W. Jaoko, K. Mandaliya, R. S. McClelland, B. A. Richardson, and J. Overbaugh. HIV-1 superinfection occurs less frequently than initial infection in a cohort of high-risk Kenyan women. *PLOS Pathogens*, 9(8):e1003593, 2013.
- L. Etienne, B.H. Hahn, P.M. Sharp, F.A. Matsen, and M. Emerman. Gene loss and adaptation to hominoids underlie the ancient origin of HIV-1. *Cell Host & Microbe*, 14(1):85–92, 2013.
- C.O. McCoy and F.A. Matsen. Abundance-weighted phylogenetic diversity measures distinguish microbial community states and are robust to sampling depth. *PeerJ*, 9:e157, 2013.
- E. H. Warren, F.A. Matsen, and J. Chou. High-throughput sequencing of B- and T-lymphocyte antigen receptors in hematology. *Blood*, 2013.
- M.M. Feeroz, K. Soliven, C.T. Small, G.A. Engel, M.A. Pacheco-Delgado, J.L. Yee, X. Wang, M. Kamrul Hasan, G. Oh, K.L. Levine, S.M.R. Alam, K. L. Craig, D.L. Jackson, E. Lee, P.A. Barry, N.W. Lerche, A.A. Escalante, F.A. Matsen, M.L. Linial, and L. Jones-Engel. Population dynamics of rhesus macaques and associated foamy virus in Bangladesh. *Emerging Microbes and Infections*, 2013.
- D.A. Nipperess and F.A. Matsen. The mean and variance of phylogenetic diversity under rarefaction. *Methods in Ecology and Evolution*, 4(6):566–572, 2013.
- F.A. Matsen, A. Gallagher, and C. McCoy. Minimizing the average distance to a closest leaf in a phylogenetic tree. *Systematic Biology*, 62(6):824–836, November 2013.
- S. J. Salipante, D. J. Sengupta, C. Rosenthal, G. Costa, J. Spangler, E. H. Sims, M. A. Jacobs, S. I. Miller, D. R. Hoogestraat, B. T. Cookson, C. O. McCoy, F. A. Matsen, J. Shendure, C. C. Lee, T. T. Harkins, and N. G. Hoffman. Rapid 16S rRNA next-generation sequencing of polymicrobial clinical samples for diagnosis of complex bacterial infections. *PLOS ONE*, 8(5):e65226, 2013.
- F.A. Matsen and S.N. Evans. Edge principal components and squash clustering: using the special structure of phylogenetic placement data for sample comparison. *PLOS ONE*, 2012.
- F.A. Matsen, A. Gallagher, and C. McCoy. nestly— a framework for running software with nested parameter choices and aggregating results. *Bioinformatics*, 29(3):387–388, 2013.
- M.T. Levine, C. McCoy, D. Vermaak, Y.C.G. Lee, M.A. Hiatt, F.A. Matsen, and H.S. Malik. Phylogenomic analysis reveals dynamic evolutionary history of the *Drosophila* heterochromatin protein 1 (HP1) gene family. *PLOS Genetics*, 8(6):e1002729, 2012.
- S. Srinivasan, N.G. Hoffman, M.T. Morgan, F.A. Matsen, T.L. Fiedler, R.W. Hall, F.J. Ross, C.O. McCoy, R. Bumgarner, J.M. Marrazzo, and D.N. Fredricks. Bacterial communities in women with bacterial vaginosis: high resolution phylogenetic analyses reveal relationships of microbiota to clinical criteria. *PLOS ONE*, 7(6):e37818, 2012.
- D.A. Lehman, D.C. Wamalwa, C.O. McCoy, F.A. Matsen, A. Langat, B.H. Chohan, S. Benki-Nugent, R. Custers-Allen, F.D. Bushman, G.C. John-Stewart, and J. Overbaugh. Low-frequency nevirapine resistance at multiple sites may predict treatment failure in infants on nevirapine-based treatment. *Journal of Acquired Immune Deficiency Syndromes*, 60(3):225, 2012.
- S. N. Evans and F. A. Matsen. The phylogenetic Kantorovich-Rubinstein metric for environmental sequence samples. *Journal of the Royal Statistical Society (B)*, 74(3):569–592, 2012.
- F.A. Matsen, N.G. Hoffman, A. Gallagher, and A. Stamatakis. A format for phylogenetic placements. *PLOS ONE*, 7(2):e31009, 2012.

- E.S. Lim, O.I. Fregoso, C.O. McCoy, F.A. Matsen, H.S. Malik, and M. Emerman. The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host and Microbe*, 11(2):194–204, 2012.
- F. A. Matsen and A. Gallagher. Reconciling taxonomy and phylogenetic inference: formalism and algorithms for describing discord and inferring taxonomic roots. *Algorithms for Molecular Biology*, 7:8, 2012.
- F.A. Matsen and S.N. Evans. Ubiquity of synonymity: almost all large binary trees are not uniquely identified by their spectra or their immanantal polynomials. *Algorithms for Molecular Biology*, 7:14, 2012.
- F.A. Matsen. constNJ: an algorithm to reconstruct sets of phylogenetic trees satisfying pairwise topological constraints. *Journal of Computational Biology*, 17(6):799–818, 2010.
- M.A. Cueto and F.A. Matsen. Polyhedral geometry of phylogenetic rogue taxa. *Bulletin of Mathematical Biology*, 73(6):1202–1226, 2011.
- F.A. Matsen, R. Kodner, and E.V. Armbrust. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1):538, 2010.
- F.A. Matsen. Fourier transform inequalities for phylogenetic trees. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 6(1):89–95, 2009.
- D. Ford, F.A. Matsen, and T. Stadler. A method for investigating relative timing information on phylogenetic trees. *Systematic Biology*, 58(2):167–183, 2009.
- F.A. Matsen and S.N. Evans. To what extent does genealogical ancestry imply genetic ancestry? *Theoretical Population Biology*, 74(2):182–190, 2008.
- F.A. Matsen, E. Mossel, and M. Steel. Mixed-up trees: the structure of phylogenetic mixtures. *Bulletin of Mathematical Biology*, 70(4):1115–1139, 2008.
- F. Matsen. Optimization over a class of tree shape statistics. *IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB)*, 4(3):506–512, 2007.
- F.A. Matsen and M. Steel. Phylogenetic mixtures on a single tree can mimic a tree of another topology. *Systematic Biology*, 56(5):767–775, 2007.
- M. Steel and F.A. Matsen. The Bayesian “star paradox” persists for long finite sequences. *Molecular Biology and Evolution*, 24(4):1075–1079, 2007.
- F.A. Matsen and J. Wakeley. Convergence to the island-model coalescent process in populations with restricted migration. *Genetics*, 172(1):701, 2006.
- F.A. Matsen. A geometric approach to tree shape statistics. *Systematic Biology*, 55(4):652–661, 2006.
- F.A. Matsen and M.A. Nowak. Win–stay, lose–shift in language learning from peers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(52):18053, 2004.