Copenhagen Phylolinguistics Workshop Syllabus

David Goldstein

3, 5, 7 May 2021

Below you will find readings that provide background to the major topics that I will cover during the workshop. Some of the readings are tutorials that are designed to introduce readers to a particular topic, whereas others assume much more knowledge. None of the readings are required for participation in the workshop. They are listed here for further edification.

Please make sure that you have downloaded the files for the course before the first class, since we will go over these during the workshop. The files can be downloaded **here**.

Day 1 (3 May)

One day one, I will go over the dataset (i.e., the Nexus file), the basics of Bayesian inference, transition models, and one of the `.Rev` scripts.

TOPIC	READING
Introduction to Phylolinguistics	Greenhill, Heggarty, et al. 2021
Introducing RevBayes	Höhna, Landis, Heath, et al. 2016, Höhna, Landis, and Heath 2017, Boussau 2018, Höhna 2019
Bayesian inference	McGrayne 2012, Gelman et al. 2009, McElreath 2016
Maximum likelihood	Baum and Smith 2013:238–247, Harmon 2019:117–131, esp. 126–131, Goldstein 2020:39–50
Transition models	Baum and Smith 2013:217–231, Cathcart 2018, Harmon 2018:106–116, Pupko and Mayrose 2020

Day 2 (5 May)

One day two, I will introduce Markov Chain Monte Carlo (MCMC), graphical models, and go over the results of the first analysis. In addition to the readings listed below, there are a number of tutorials on RevBayes **here**.

TOPIC	READING
Introduction to Markov Chain Monte Carlo (MCMC)	Shaver 2017, Harmon 2019:28–34, van Ravenzwaaij et al. 2018
Bayesian inference	Baum and Smith 2013:247–259, Harmon 2019:28–34, Holder and Lewis 2003, Alfaro and Holder 2006
Graphical models	Freyman n.d.

Day 3 (7 May)

One day three, I plan to cover among-site rate variation (ASRV) and model comparison. The RevBayes tutorial on ASRV can be found here and that on model comparison here.

TOPIC	READING
Rate heterogeneity	Yang 2014:118–119
Model comparison	Harmon 2019:21–23, Höhna, Landis, and Heath 2017

Bayesian phylolinguistic studies

CLADE	LITERATURE
Iranian	Cathcart 2019, Cathcart 2020
Semitic	Kitchen et al. 2009
Dravidian	Kolipakam et al. 2018
Transeurasian	Robbeets and Bouckaert 2018
Pama-Nyungan	Bowern and Atkinson 2012, Bouckaert, Bowern, et al. 2018, Bowern 2012
Turkic	Savelyev and Robbeets 2020
Sino-Tibetan	Sagart et al. 2019, Zhang, Yan, et al. 2019, Zhang, Ji, et al. 2020
Japonic	Lee and Hasegawa 2011
Austronesian	Dunn et al. 2008, Saunders 2005, Gray, Drummond, et al. 2009, Greenhill and Gray 2009, Greenhill, Atkinson, et al. 2010
Indo-European	Gray and Atkinson 2003, Atkinson and Gray 2006, Bouckaert, Lemey, et al. 2012, Chang et al. 2015, Rama 2018
Slavic	Cathcart and Wandl 2020
Dene-Yeniseian	A. Sicoli and Holton 2014, Yanovich 2020
Bantu	Guillon and Mace 2016; Holden et al. 2005

References

A. Sicoli, Mark and Gary Holton (2014). Linguistic phylogenies support back-migration from Beringia to Asia. *PLoS ONE* 9.3 (Mar. 2014), e91722. DOI: 10.1371/journal.pone.0091722. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0091722.

Alfaro, Michael E. and Mark T. Holder (2006). The posterior and the prior in Bayesian phylogenetics. *Annual Review of Ecology, Evolution, and Systematics* 37, 19–42. DOI: 10.1146/annurev.ecolsys.37.091305.110021.

Atkinson, Quentin D. and Russell D. Gray (2006). How old is the Indo-European language family? Illumination or more moths to the flame? *Phylogenetic methods and the prehistory of languages*. Ed. by Peter Forster and Colin A. Renfrew. Cambridge: McDonald Institute for Archaeological Research, 91–109.

Baum, David A. and Stacey D. Smith (2013). *Tree thinking. An introduction to phylogenetic biology*. Greenwood Village, CO: Roberts and Co.

Bouckaert, Remco R., Claire Bowern, and Quentin D. Atkinson (2018). The origin and expansion of Pama-Nyungan languages across Australia. *Nature Ecology & Evolution*. DOI: 10.1038/s41559-018-0489-3. URL: https://natureecoevocommunity.nature.com/users/31079-quentin-atkinson/posts/30473-the-origin-and-expansion-of-the-world-s-largest-hunter-gatherer-language-family.

Bouckaert, Remco R., Philippe Lemey, Michael Dunn, Simon J. Greenhill, Alexander V. Alekseyenko, Alexei J. Drummond, Russell D. Gray, Marc A. Suchard, and Quentin D. Atkinson (2012). Mapping the origins and expansion of the Indo-European language family. *Science* 337.6097 (Aug. 2012), 957–960. DOI: 10.1126/science.1219669. URL: http://science.sciencemag.org/content/337/6097/957/tab-pdf.

Boussau, Bastien (2018). RevBayes. Bayesian inference in phylogenetics using graphical models and a R-like language. June 2018. URL: http://genome.jouy.inra.fr/applibugs/applibugs.18_06_21.bboussau.pdf.

Bowern, Claire (2012). The riddle of Tasmanian languages. *Proceedings of the Royal Society B: Biological Sciences* 279.1747 (Nov. 2012), 4590–4595. DOI: 10.1098/rspb.2012.1842.

Bowern, Claire and Quentin D. Atkinson (2012). Computational phylogenetics and the internal structure of Pama-Nyungan. *Language* 88.4 (Dec. 2012), 817–845. DOI: 10.1353/lan.2012.0081. URL: https://muse.jhu.edu/article/492550.

Cathcart, Chundra Aroor (2018). Modeling linguistic evolution. A look under the hood. *Linguistics Vanguard* 4.1. DOI: 10.1515/lingvan-2017-0043.

– (2019). Dialectal layers in West Iranian. A hierarchical Dirichlet process approach to linguistic relationships. URL: https://arxiv.org/pdf/2001.05297.pdf.

– (2020). A probabilistic assessment of the Indo-Aryan Inner-Outer Hypothesis. *Journal of Historical Linguistics* 10.1, 42–86.

Cathcart, Chundra Aroor and Florian Wandl (2020). In search of isoglosses. Continuous and discrete language embeddings in Slavic historical phonology. URL: https://arxiv.org/pdf/2005.13575.pdf.

Chang, Will, Chundra Aroor Cathcart, David P. Hall, and Andrew J. Garrett (2015). Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis. *Language* 91.1 (Mar. 2015), 194–244. DOI: 10.1353/lan.2015.0005. URL: https://muse.jhu.edu/article/576998.

Dunn, Michael, Stephen C. Levinson, Eva Lindström, Ger Reesink, and Angela Terrill (2008). Structural phylogeny in historical linguistics. Methodological explorations applied in Island Melanesia. *Language* 84, 710–759.

Freyman, William A. (n.d.). *Introduction to graphical models. A gentle introduction to graphical models, probabilistic programming, and MCMC using a simple linear regression example*. URL: https://revbayes.github.io/tutorials/intro/graph_models.html.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (2009). *Bayesian data analysis*. 3rd ed. Boca Raton: CRC Press. URL: http://www.stat.columbia.edu/~gelman/book/.

Goldstein, David M. (2020). Indo-European phylogenetics with R. A tutorial introduction. *Indo-European Linguistics* 8.1, 110–180. DOI: 10.1163/22125892-20201000.

Gray, Russell D. and Quentin D. Atkinson (2003). Language-tree divergence times support the Anatolian theory of Indo-European origin. *Nature* 426, 435–439. DOI: 10.1038/nature02029. URL: https://www.nature.com/articles/nature02029.

Gray, Russell D., Alexei J. Drummond, and Simon J. Greenhill (2009). Language phylogenies reveal expansion pulses and pauses in Pacific settlement. *Science* 323.5913, 479–483. DOI: 10.1126/science.1166858. URL: https://science.sciencemag.org/content/323/5913/479.

Greenhill, Simon J., Quentin D. Atkinson, Andrew Meade, and Russell D. Gray (2010). The shape and tempo of language evolution. Biological Sciences. *Proceedings of the Royal Society B* 277 (Aug. 2010), 2443–2450. DOI: 10.1098/rspb.2010.0051. URL: http://rspb.royalsocietypublishing.org/content/277/1693/2443.

Greenhill, Simon J. and Russell D. Gray (2009). Austronesian language phylogenies. Myths and misconceptions about Bayesian computational methods. *Austronesian historical linguistics and culture history. A festschrift for Robert Blust*. Ed. by Alexander Adelaar and Andrew Pawley. Canberra: Pacific Linguistics, 375–397. URL: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.799.8319&rep=rep1&type=pdf.

Greenhill, Simon J., Paul Heggarty, and Russell D. Gray (2021). Bayesian phylolinguistics. *The handbook of historical linguistics*. Ed. by Brian D. Joseph, Richard D. Janda, and Barbara S. Vance. Vol. 2. Malden, MA: Wiley-Blackwell. DOI: 10.1002/9781118732168.ch11.

Guillon, Myrtille and Ruth Mace (2016). A phylogenetic comparative study of Bantu kinship terminology finds limited support for its co-evolution with social organisation. *PLoS ONE* 11.5 (Mar. 2016), e0155170. DOI: 10.1371/journal.pone.0147920. URL: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0147920.

Harmon, Luke J. (2018). Phylogenetic comparative methods. Models of discrete character evolution. Mar. 2018.

– (2019). Phylogenetic comparative methods. Learning from trees. Mar. 2019. URL: https://lukejharmon.github.io/pcm/.

Höhna, Sebastian (2019). *Getting started with RevBayes. A very basic overview on how to use RevBayes*. Apr. 2019. URL: https://revbayes.github.io/tutorials/intro/#overview.

Höhna, Sebastian, Michael J. Landis, and Tracy A. Heath (2017). Phylogenetic inference using `RevBayes`. *Current Protocols in Bioinformatics* 57.1, 6.16.1–6.16.34. DOI: 10 . 1002 / cpbi . 22. URL: https://currentprotocols.onlinelibrary.wiley.com/doi/full/10.1002/cpbi.22.

Höhna, Sebastian, Michael J. Landis, Tracy A. Heath, Bastien Boussau, Nicolas Lartillot, Brian R. Moore, John P. Huelsenbeck, and Fredrik Ronquist (2016). RevBayes. Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology* 65.4 (July 2016), 726–736. DOI: 10.1093/sysbio/syw021. URL: https://academic.oup.com/sysbio/article/65/4/726/1753608.

Holden, Clare Janaki, Andrew Meade, and Mark Pagel (2005). Comparison of maximum parsimony and Bayesian Bantu language trees. A phylogenetic approach. *The evolution of cultural diversity*. Ed. by Ruth Mace, Clare Janaki Holden, and Stephen J. Shennan. Walnut Creek: Left Coast Press, 53–66. URL: https://www.taylorfrancis.com/books/e/9781315418612/chapters/10.4324/9781315418612-12.

Holder, Mark T. and Paul O. Lewis (2003). Phylogeny estimation. Traditional and Bayesian approaches. *Nature Reviews Genetics* 4.4 (May 2003), 275–284. DOI: 10.1038/nrg1044.

Kitchen, Andrew, Christopher Ehret, Shiferaw Assefa, and Connie J. Mulligan (2009). Bayesian phylogenetic analysis of Semitic languages identifies an Early Bronze Age origin of Semitic in the Near East. *Proceedings of the Royal Society B: Biological Sciences* 276.1668 (Apr. 2009). DOI: 10.1098/rspb.2009.0408. URL: https://royalsocietypublishing.org/doi/full/10.1098/rspb.2009.0408.

Kolipakam, Vishnupriya, Fiona M. Jordan, Michael Dunn, Simon J. Greenhill, Remco R. Bouckaert, Russell D. Gray, and Annemarie Verkerk (2018). A Bayesian phylogenetic study of the Dravidian language family. *Royal Society Open Science* 5 (Mar. 2018). DOI: 10.1098/rsos.171504. URL: http://rsos.royalsocietypublishing.org/content/5/3/171504.

Lee, Sean and Toshikazu Hasegawa (2011). Bayesian phylogenetic analysis supports an agricultural origin of Japonic languages. Biological Sciences. *Proceedings of the Royal Society B: Biological Sciences* 278.1725, 3662–3669. DOI: 10.1098/rspb.2011.0518. URL: https://royalsocietypublishing.org/doi/full/10.1098/rspb.2011.0518.

McElreath, Richard (2016). *Statistical rethinking*. 2nd ed. Boca Raton: CRC Press. URL: https://xcelab.net/rm/statistical-rethinking/.

McGrayne, Sharon Bertsch (2012). *The theory that would not die. How Bayes' rule cracked the Enigma code, hunted down Russian submarines, and emerged triumphant from two centuries of controversy*. New Haven: Yale University Press.

Pupko, Tal and Itay Mayrose (2020). A gentle introduction to probabilistic evolutionary models. *Phylogenetics in the genomic era*. Ed. by Celine Scornavacca, Frédéric Delsuc, and Nicolas Galtier. Chap. 1.1, 1–21.

Rama, Taraka (2018). Three tree priors and five datasets. A study of the effect of tree priors in Indo-European phylogenetics. *Language Dynamics and Change* 8.2, 182–218. DOI: 10.1163/22105832-00802005. URL: https://arxiv.org/abs/1805.03645.

Robbeets, Martine and Remco R. Bouckaert (2018). Bayesian phylolinguistics reveals the internal structure of the Transeurasian family. *Journal of Language Evolution* 3.2 (July 2018), 145–162. DOI: 10.1093/jole/lzy007. URL: https://academic.oup.com/jole/article/3/2/145/5067185.

Sagart, Laurent, Guillaume Jacques, Yunfan Lai, Robin J. Ryder, Valentin Thouzeau, Simon J. Greenhill, and Johann-Mattis List (2019). Dated language phylogenies shed light on the ancestry of Sino-Tibetan. *Proceedings of the National Academy of Sciences of the United States of America*. DOI: 10.1073/pnas.1817972116. URL: https://www.shh.mpg.de/1285923/sino-tibetan-origin.

Saunders, Arpiar (2005). Linguistic phylogenetics of the Austronesian family. A performance review of methods adapted from biology. B.A. thesis. URL: https://www.swarthmore.edu/sites/default/files/assets/documents/linguistics/2006_saunders_arpy.pdf.

Savelyev, Alexander and Martine Robbeets (2020). Bayesian phylolinguistics infers the internal structure and the time-depth of the Turkic language family. *Journal of Language Evolution* 5.1 (Jan. 2020), 39–53. DOI: 10.1093/jole/lzz010. URL: https://academic.oup.com/jole/article/5/1/39/5736268?searchresult=1.

Shaver, Ben (2017). *A zero-math introduction to Markov Chain Monte Carlo methods*. URL: https://towardsdatascience.com/a-zero-math-introduction-to-markov-chain-monte-carlo-methods-dcba889e0c50.

Van Ravenzwaaij, Don, Pete Cassey, and Scott D. Brown (2018). A simple introduction to Markov Chain Monte-Carlo sampling. *Psychonomic Bulletin & Review* 25, 143–154. DOI: 10.3758/s13423-016-1015-8. URL: https://link.springer.com/article/10.3758/s13423-016-1015-8#citeas.

Yang, Ziheng (2014). *Molecular evolution. A statistical approach*. Oxford: Oxford University Press.

Yanovich, Igor (2020). Phylogenetic linguistic evidence and the Dene-Yeniseian homeland. *Diachronica* 37.3 (Apr. 2020), 410–446.

Zhang, Hanzhi, Ting Ji, Mark Pagel, and Ruth Mace (2020). Dated phylogeny suggests early Neolithic origin of Sino-Tibetan languages. *Scientific Reports* 10.20792 (Nov. 2020). DOI: 10.1038/s41598-020-77404-4. URL: https://www.nature.com/articles/s41598-020-77404-4#citeas.

Zhang, Menghan, Shi Yan, Wuyun Pan, and Li Jin (2019). Phylogenetic evidence for Sino-Tibetan origin in northern China in the Late Neolithic. *Nature* 569, 112–115. DOI: 10.1038/s41586-019-1153-z. URL: https://www.nature.com/articles/s41586-019-1153-z.