University of Groningen

What fruits can we get from this tree?

Laudanno, Giovanni

DOI:
[10.33612/diss.155031292](https://doi.org/10.33612/diss.155031292)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Laudanno, G. (2021). *What fruits can we get from this tree? A journey in phylogenetic inference through likelihood modeling*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.
<https://doi.org/10.33612/diss.155031292>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

rijksuniversiteit
 groningen

What fruits can we get from this tree?

A journey in phylogenetic inference through likelihood modeling

PhD thesis

to obtain the degree of PhD at the
University of Groningen
on the authority of the
Rector Magnificus Prof. C. Wijmenga
and in accordance with
the decision by the College of Deans.

This thesis will be defended in public on

Monday 11 January 2021 at 16.15 hours

by

Giovanni Laudanno

born on 11 January 1987
in Napoli, Italy

Supervisor

Prof. R. S. Etienne

Co-supervisor

Dr. B. Haegeman

Assessment Committee

Prof. E. C. Wit

Prof. F. Hartig

Prof. L. J. Harmon

Every path is the right path. Everything could've been anything else. And it would have just as much meaning.

– Mr. Nobody

E allora capii, fui costretto a capire, che fare il dottore è soltanto un mestiere, che la scienza non puoi regalarla alla gente, se non vuoi ammalarti dell'identico male.

– Fabrizio De André

Cover design by Tatyana Zabanova.

Cover editing by Giovanni Laudanno and Giuseppe Marzano.

Contents

	Page
1. Introduction	1
1.1. Biological background	3
1.1.1. Evolution	3
1.1.2. Speciation	4
1.2. Phylogenetic Inference	5
1.2.1. The Bayesian framework	6
1.2.2. Bayesian framework in phylogenetics	7
1.2.3. The likelihood	8
1.2.4. The tree prior	10
1.3. Likelihood modeling	10
1.3.1. Complete tree vs reconstructed tree	11
1.3.2. The Birth-Death model	11
1.3.3. The Q-framework	13
1.3.4. Simulating phylogenies	14
1.3.5. Maximum Likelihood Estimation	15
1.4. Thesis outline	16
2. Additional analytical support for a new method to compute the likelihood of diversification models	17
2.1. Introduction	19
2.2. The diversity-dependent diversification model	21
2.3. The Q-framework	25
2.4. The likelihood for the diversity-independent case	28
2.5. Equivalence for the diversity-independent case	31

2.6.	A note on sampling a fraction of extant species	36
2.7.	The diversity-dependent case without extinction	40
2.8.	Concluding remarks	42

3. Exploring the multiple birth hypothesis: a likelihood for crowded phylogenies 45

3.1.	Introduction	47
3.2.	Methods	48
3.2.1.	Model	48
3.2.2.	Mathematical description: the P-framework	50
3.2.3.	Simulations	51
3.2.4.	Likelihood derivation: the Q-framework	52
3.2.5.	Conditional probabilities	56
3.2.6.	Results	58
3.2.7.	Detecting the MBD signal in phylogenies	61
3.3.	Discussion	68
3.4.	mbd code	69
3.5.	Acknowledgements	69
3.6.	Appendix A: Additional statistics	71
3.7.	Appendix B: Q-approach conditioning	73

4. Detecting lineage-specific shifts in diversification: a proper likelihood approach 75

4.1.	Introduction	77
4.2.	Methods	78
4.2.1.	The D-E framework	78
4.2.2.	The D-E framework applied to mapped rate shifts leads to probabilities larger than 1	79
4.2.3.	Corrected likelihood - Example	82
4.2.4.	Corrected likelihood - General case	84
4.2.5.	Conditional likelihoods	86
4.2.6.	Performance of the corrected likelihood in parameter estimation	87
4.2.7.	Extending the corrected likelihood $\mathcal{L}_{\text{corr}}$ to time-dependent and diversity-dependent diversification rates	89
4.2.8.	Extending the corrected likelihood $\mathcal{L}_{\text{corr}}$ to multiple shifts	89
4.2.9.	Detecting rate shifts in phylogenetic trees	90
4.3.	Discussion	91

4.4.	Appendix A: D-E likelihood for example phylogeny of Fig. 4.1 . . .	96
4.5.	Appendix B: Corrected likelihood for phylogeny of Fig. 4.1	98
4.5.1.	A short introduction to Nee et al.	98
4.5.2.	Comparison of D-E likelihood and corrected likelihood . . .	99
4.6.	Appendix C: Corrected likelihood for general phylogenies	102
4.6.1.	A useful identity	102
4.6.2.	Case without rate shift	103
4.6.3.	Case of a single rate shift	106
4.7.	Appendix D: Likelihood for unobserved rate shift	111
4.8.	Appendix E: Rate shifts in diversity-dependent model	113
5.	Quantifying the impact of an inference model in Bayesian	
	phylogenetics	117
5.1.	Introduction	119
5.2.	Description	121
5.2.1.	pirouette's pipeline	122
5.2.2.	Controls	126
5.3.	Usage	127
5.4.	Discussion	130
5.5.	Acknowledgments	132
5.6.	Authors' contributions	132
5.7.	Data accessibility	132
5.8.	Supplementary material	133
5.8.1.	Guidelines for users	134
5.8.2.	Installation	134
5.8.3.	Resources	135
5.8.4.	Citation of pirouette	136
5.8.5.	The twinning process	136
5.8.6.	Candidate models	137
5.8.7.	Stochasticity caused by simulating phylogenies	138
5.8.8.	The nLTT statistic	138
5.8.9.	Main functions	138
5.8.10.	Main example	139
5.8.11.	Using a distribution of trees	146
5.8.12.	The effect of the number of taxa	148
5.8.13.	The effect of DNA sequence length	153
5.8.14.	The effect of assuming a Yule tree prior on a Yule tree . . .	157
5.8.15.	The effect of assuming a Yule tree prior on a BD tree . . .	159

5.8.16. The effect of DD trees at different degrees of likelihood .	161
5.8.17. The effect of equal or equalized mutation rate in the twin alignment	164
5.8.18. The effect of mutation rate	166
6. Synthesis	175
6.1. Modelling biology	177
6.2. Different paradigms	179
6.3. Inference limitations in Birth Death models	179
6.3.1. Specific limitations of the Q-framework	181
6.4. Future prospects	182
6.4.1. Applications of chapter 2	182
6.4.2. Applications of chapter 3	184
6.4.3. Applications of chapter 4	187
6.4.4. Applications of chapter 5	188
Bibliography	191
7. Summary	203
8. Samenvatting	207
Curriculum Vitae	211