

Inferring Phylogenetic Networks Using PhyloNet

DINGQIAO WEN¹, YUN YU¹, JIAFAN ZHU¹, LUAY NAKHLEH^{1,2,*}

¹Computer Science, Rice University, Houston, TX, USA; ²BioSciences, Rice University, Houston, TX, USA;

*Correspondence to be sent to: Computer Science, Rice University, Houston, TX, USA; E-mail: nakhleh@rice.edu.

Abstract.—PhyloNet was released in 2008 as a software package for representing and analyzing phylogenetic networks. At the time of its release, the main functionalities in PhyloNet consisted of measures for comparing network topologies and a single heuristic for reconciling gene trees with a species tree. Since then, PhyloNet has grown significantly. The software package now includes a wide array of methods for inferring phylogenetic networks from data sets of unlinked loci while accounting for both reticulation (e.g., hybridization) and incomplete lineage sorting. In particular, PhyloNet now allows for maximum parsimony, maximum likelihood, and Bayesian inference of phylogenetic networks from gene tree estimates. Furthermore, Bayesian inference directly from sequence data (sequence alignments or bi-allelic markers) is implemented. Maximum parsimony is based on an extension of the “minimizing deep coalescences” criterion to phylogenetic networks, whereas maximum likelihood and Bayesian inference are based on the multispecies network coalescent. All methods allow for multiple individuals per species. As computing the likelihood of a phylogenetic network is computationally hard, PhyloNet allows for evaluation and inference of networks using a pseudo-likelihood measure. PhyloNet summarizes the results of the various analyses, and generates phylogenetic networks in the extended Newick format that is readily viewable by existing visualization software. [phylogenetic networks; reticulation; incomplete lineage sorting; multispecies network coalescent; Bayesian inference; maximum likelihood; maximum parsimony.]

1 With the increasing availability of whole-genome and
2 multi-locus data sets, an explosion in the development of
3 methods for species tree inference from such data ensued.
4 In particular, the multispecies coalescent (Degnan and
5 Rosenberg, 2009) played a central role in explaining and
6 modeling the phenomenon of gene tree incongruence due
7 to incomplete lineage sorting (ILS), as well as in devising
8 computational methods for species tree inference in the
9 presence of ILS; e.g., (Heled and Drummond, 2010; Liu,
10 2008).

11 Nevertheless, with the increasing recognition that the
12 evolutionary histories of several groups of closely related
13 species are reticulate (Mallet et al., 2016), there is need
14 for developing methods that infer species phylogenies
15 while accounting not only for ILS but also for processes
16 such as hybridization. Such reticulate species phylogenies
17 are modeled by *phylogenetic networks* (Nakhleh, 2010).
18 A phylogenetic network extends the phylogenetic tree
19 model by allowing for horizontal edges that capture the
20 inheritance of genetic material through gene flow (Fig.
1(a)). While the phylogenetic network captures how the

23 within its branches capture the evolutionary histories
24 of individual, recombination-free loci (Fig. 1(b)). The
25 relationship between phylogenetic networks and trees
26 is complex in the presence of ILS (Zhu et al., 2016).
27 Mathematically, the topology of a phylogenetic network
28 takes the form of a rooted, directed, acyclic graph. In
29 particular, while gene flow involves contemporaneous
30 species or populations, past extinctions or incomplete
31 sampling for taxa sometimes result in horizontal edges
32 that appear to be “forward in time” (Fig. 1). It is
33 important to account for such an event, which is why
34 acyclicity, rather than having truly horizontal edges, is
35 the only constraint that should be imposed on rooted
36 directed graphs, in practice, if one is to model reticulate
37 evolutionary histories.

38 For inference of phylogenetic networks from multi-
39 locus data sets, the notions of coalescent histories and
40 the multispecies coalescent were extended to
41 phylogenetic networks (Yu et al., 2012, 2011). Based on
42 these new models, the “minimizing deep coalescence”
43 criterion (Maddison, 1997; Than and Nakhleh, 2009) was
44 extended to phylogenetic networks, which allowed for a
45 maximum parsimony inference of phylogenetic networks
46 from the gene tree estimates of unlinked loci (Yu et al.,
47 2013a). Subsequently, maximum likelihood inference
48 (from gene tree estimates) via hill-climbing heuristics
49 and Bayesian inference via reversible-jump Markov chain
50 Monte Carlo (RJMCMC) were devised (Wen et al.,
51 2016; Yu et al., 2014). As computing the likelihood of a
52 phylogenetic network formed a major bottleneck in the
53 inference, speedup techniques for likelihood calculations
54 and pseudo-likelihood of phylogenetic networks were
55 introduced (Yu and Nakhleh, 2015b; Yu et al., 2013b).
56 Finally, to enable direct estimation from sequence data,
57 new methods were developed for Bayesian inference
58 from sequence alignments of unlinked loci (Wen and
59 Nakhleh, 2017) as well as bi-allelic markers of unlinked

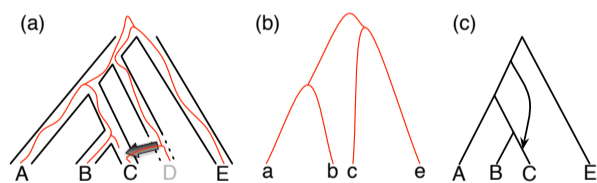


FIGURE 1. (a) A phylogenetic network on five taxa, with taxon D missing (due to extinction or incomplete sampling), and a hybridization involving (ancestors of) taxa D and C . Shown within the branches of the phylogenetic network is a tree of a recombination-free locus whose evolutionary history includes introgression. (b) The gene tree that would be estimated, barring inference error, on the locus illustrated in (a). (c) An abstract depiction of the phylogenetic network of (a) given that taxon D is missing.

21 species, or populations, have evolved, gene trees growing
22

1 loci (Zhu et al., 2017). Here we introduce PhyloNet 3, a
 2 software package for phylogenetic network inference from
 3 multi-locus data under the aforementioned models and
 4 criteria. This version is a significant expansion of the
 5 version reported on in (Than et al., 2008). Phylogenetic
 6 networks inferred by PhyloNet are represented using an
 7 extended Newick format and can be readily visualized by
 8 Dendroscope (Huson and Scornavacca, 2012).

9 MODELS AND MAIN INFERENCE FEATURES

10 *Simple Counts of Extra Lineages: Maximum Parsimony*

11 Minimizing the number of deep coalescences, or
 12 MDC, is a criterion that was proposed originally by
 13 Maddison (Maddison, 1997) for species tree inference
 14 and later implemented and tested in both heuristic form
 15 (Maddison and Knowles, 2006) and exact algorithms
 16 (Than and Nakhleh, 2009). Yu et al. (Yu et al., 2013a)
 17 extended the MDC criterion to phylogenetic networks.
 18 The **InferNetwork_MP** command infers a species
 19 network with a specified number of reticulation nodes
 20 under the extended MDC criterion. Inference under this
 21 criterion is done via a local search heuristic, and the
 22 phylogenetic networks returned by the program include,
 23 in addition to the topologies, the inheritance probability
 24 estimates, as well as the number of extra lineages on
 25 each branch of the network, and the total number of
 26 extra lineages of the phylogenetic network. For this
 27 program, only gene tree topologies are used as input
 28 (that is, gene tree branch lengths are irrelevant), and
 29 the number of individuals per species could vary across
 30 loci. Furthermore, to account for uncertainty in the input
 31 gene tree estimates, the program allows for a set of
 32 gene trees per locus that could be obtained from a
 33 bootstrap analysis or a posterior sample on the sequences
 34 of the respective locus. For inference under the MDC
 35 criterion, the maximum number of reticulation events in
 36 the phylogenetic network must be specified *a priori*. Full
 37 details of the MDC criterion for phylogenetic networks
 38 and the inference heuristics can be found in (Yu et al.,
 39 2013a).

40 Inference based on the MDC criterion does not allow
 41 for estimating branch lengths or any other associated
 42 parameters of the inferred phylogenetic network beyond
 43 the topology and inheritance probabilities.

44 *When the Species Phylogeny’s Branches Are Too Short: 45 Maximum Likelihood*

46 One limitation of inference based on the MDC criterion
 47 is the inability to estimate parameter values beyond
 48 the network’s topology. Another limitation is the fact
 49 that such inference is not statistically consistent for
 50 species trees (Than and Rosenberg, 2011), which implies
 51 problems in the case of phylogenetic network inference
 52 based on the criterion as well (more on the notion of
 53 “statistical consistency” in the case of networks below).
 54 The latter problem arises especially when the species

55 phylogeny has very short branches. To address these
 56 two limitations, Yu et al. (Yu et al., 2014) implemented
 57 maximum likelihood estimation of phylogenetic networks
 58 based on the multispecies network coalescent (Yu
 59 et al., 2012). The **InferNetwork_ML** command infers
 60 a maximum likelihood species network(s) along with
 61 its branch lengths (in coalescent units) and inheritance
 62 probabilities. During the search, the branch lengths and
 63 inheritance probabilities of a proposed species network
 64 can be either sampled or optimized (the former is much
 65 faster and has been shown to perform very well). The
 66 input consists of either rooted gene tree topologies
 67 alone, or rooted gene trees with branch lengths (in
 68 coalescent units). If the gene tree branch lengths are
 69 to be used, the gene trees must be ultrametric. As in
 70 the case of maximum parsimony inference, local search
 71 heuristics are used to obtain the maximum likelihood
 72 estimates. Furthermore, multiple individuals per species
 73 could be used, and their numbers could vary across loci.
 74 Multiple gene trees per locus could be used, as above,
 75 to account for uncertainty in the gene tree estimates.
 76 The user can either specify the maximum number of
 77 reticulation events *a priori* or utilize the cross-validation
 78 (the **InferNetwork_ML_CV** command) or bootstrap
 79 (the **InferNetwork_ML_Bootstrap** command) to
 80 determine the model complexity. Furthermore, several
 81 information criteria (AIC, BIC, and AICs) are
 82 implemented. Full details of the maximum likelihood
 83 inference of phylogenetic networks and the inference
 84 heuristics can be found in (Yu et al., 2014).

85 It is important to note that computing the
 86 likelihood of a phylogenetic network is a major
 87 computational bottleneck in all statistical inference
 88 methods implemented in PhyloNet. To ameliorate
 89 this problem, PhyloNet also allows for inference of
 90 phylogenetic networks based on a “pseudo-likelihood”
 91 measure, via the **InferNetwork_MPL** command.
 92 However, for this method, the input could consist
 93 only of gene tree topologies (branch lengths are not
 94 allowed). Multiple individuals per species, as well as non-
 95 binary gene trees, are also allowed. Full details about
 96 inference under pseudo-likelihood can be found in (Yu
 97 and Nakhleh, 2015b).

98 *Penalizing Network Complexity: Bayesian Inference*

99 Discussing statistical inference in general, Attias
 100 (Attias, 1999) listed three problems with maximum
 101 likelihood: “First, it produces a model that overfits
 102 the data and subsequently have [sic] suboptimal
 103 generalization performance. Second, it cannot be used to
 104 learn the structure of the graph, since more complicated
 105 graphs assign a higher likelihood to the data. Third, it
 106 is computationally tractable only for a small class of
 107 models.” When the model of interest in a phylogenetic
 108 tree, the first two problems are generally not of concern
 109 (barring the complexity of the model of evolution
 110 underlying the inference). Putting aside the problem of
 111 computational tractability, the first two problems listed

1 by Attias are an Achilles heel for phylogenetic network
2 inference by maximum likelihood.

3 Phylogenetic networks can be viewed as mixture
4 models whose components are distributions defined
5 by parental trees of the network (Zhu et al., 2016).
6 Inferring the true model in the case of phylogenetic
7 networks includes determining, in addition to many
8 other parameters, the true number of reticulations.
9 Inference of such a model based on an unpenalized
10 likelihood can rarely work since, as Attias pointed out,
11 “more complicated graphs assign a higher likelihood.”
12 In fact, the notion of *statistical consistency*, which
13 has been a staple in the literature on phylogenetic
14 tree inference, is not even applicable in the case
15 of maximum (unpenalized) likelihood of phylogenetic
16 networks—it is easy to imagine scenarios where adding
17 more reticulations to the true phylogeny would only
18 improve the likelihood. Therefore, in analyzing data
19 sets under the aforementioned likelihood-based inference
20 methods we recommend experimenting with varying the
21 maximum allowed number of reticulation events and
22 inspecting the likelihoods of the models with different
23 numbers of reticulations.

24 A more principled way to deal with model complexity
25 is via Bayesian inference, which allows, among other
26 things, for regularization via the prior distribution. The
27 **MCMC_GT** command performs Bayesian inference
28 of the posterior distribution of the species network
29 along with its branch lengths (in coalescent units)
30 and inheritance probabilities via reversible-jump Markov
31 chain Monte Carlo (RJMCMC). The input consists of
32 single or multiple rooted gene tree topologies per locus,
33 as above, and the number of individuals per species could
34 vary across loci. Full details of the Bayesian inference of
35 phylogenetic networks can be found in (Wen et al., 2016).

36 To handle gene tree uncertainty in a principled
37 manner, and to allow for inferring the values of various
38 network-associated parameters, PhyloNet implements
39 Bayesian inference of phylogenetic networks directly
40 from sequence data. The **MCMC_SEQ** command
41 performs Bayesian inference of the posterior distribution
42 of the species network along with its divergence
43 times (in units of expected number of mutations
44 per site) and population mutation rates (in units of
45 population mutation per site), inheritance probabilities,
46 and ultrametric gene trees along with its coalescent
47 times (in units of expected number of mutations
48 per site) simultaneously via RJMCMC. The input
49 consists of sequence alignments of unlinked loci. Multiple
50 individuals per species could be used, and their
51 numbers could vary across loci. Full details of the
52 co-estimation method can be found in (Wen and
53 Nakhleh, 2017). The **MCMC_BiMarkers** command,
54 on the other hand, performs Bayesian inference of
55 the posterior distribution of the species network
56 along with its divergence times (in units of expected
57 number of mutations per site), population mutation
58 rates (in units of population mutation per site), and
59 inheritance probabilities via RJMCMC. The input
60 consists of bi-allelic markers of unlinked loci, most

61 notably single nucleotide polymorphisms (SNPs) and
62 amplified fragment length polymorphisms (AFLP). Also,
63 multiple individuals per species could be used. This
64 method carries out numerical integration over all gene
65 trees, which allows it to completely sidestep the issue of
66 sampling gene trees. Full details of the computation can
67 be found in (Zhu et al., 2017).

68 *Other Features*

69 In addition to the aforementioned inference methods
70 and all the functionalities that existed in the old version
71 of PhyloNet (Than et al., 2008), the software package
72 includes new features that help with other types of
73 analyses.

74 As trees are a special case of networks, all the features
75 above allow for species tree inference by simply setting
76 the number of reticulations allowed during the analysis to
77 0. Additionally, PhyloNet implements greedy consensus,
78 the “democratic vote,” and the GLASS method of
79 (Mossel and Roch, 2010).

80 PhyloNet also includes a method for distance-based
81 inference of phylogenetic networks (Yu and Nakhleh,
82 2015a), as well as a Gibbs sampling method for
83 estimating the parameters of a given phylogenetic
84 network (Yu et al., 2016).

85 Last but not least, the **SimGTinNetwork** and
86 **SimBiMarkersinNetwork** simulate gene trees and bi-
87 allelic markers, respectively, on a phylogenetic network.
88 In particular, the former automates the process of
89 simulating gene trees in the presence of reticulation
90 and incomplete lineage sorting using the program
91 **ms** (Hudson, 2002). The latter command extends the
92 simulator developed in (Bryant et al., 2012).

93 INPUT AND OUTPUT FORMATS

94 PhyloNet 3 is a software package in the JAR format
95 that can be installed and executed on any system with
96 the Java Platform (Version 7.0 or higher). The command
97 line in a command prompt is

```
98 java -jar PhyloNet_X.Y.Z.jar script.nex
```

99 where **X.Y.Z** is the version number (version 3.6.1
100 is the most recent release), and **script.nex** is the
101 input NEXUS file containing data and the PhyloNet
102 commands to be executed.

103 The input data and the commands are listed in
104 blocks. Each block start with the “BEGIN” keyword and
105 terminate with the “END;” keyword. Commands in a
106 PHYLONET block begin with a command identifier and
107 terminate with a semicolon.

108 In the example input file below, to estimate the
109 posterior distribution of the species network and the gene
110 trees, the input sequence alignments are listed in the
111 “DATA” block. The starting gene tree for each locus
112 and the starting network, which is optional, can be
113 specified in the “TREES” block and the “NETWORKS”

1 block, respectively. Finally the command *MCMC_SEQ*
 2 and its parameters are provided for execution in the
 3 “PHYLONET” block. Details about specific parameters
 4 for a given command can be found on the website of
 5 PhyloNet.

```

6 #NEXUS
7
8 BEGIN DATA;
9     Dimensions ntax=3 nchar=35;
10    Format datatype=dna symbols="ACTG" missing=?
11    gap=-;
12    Matrix
13    [locus0, 15]
14    A TCGCGCTAACGTCCA
15    B GCGCACCTACTGCGG
16    B GCGCACCTACTCCGG
17    [locus1, 20]
18    A GAAACGGATCTAAGTGTACG
19    B CGCTCGGATCTAAGTGTACG
20    C CGCTCGGATCTAAGTGTACG
21    ;END;
22
23 BEGIN TREES;
24 Tree gt0 = (A:0.119,(B:0.058,C:0.058):0.061);
25 Tree gt1 = (A:0.068,(B:0.016,C:0.016):0.052);
26 END;
27
28 BEGIN NETWORKS;
29 Network net = ((B:0.0)#H1:0.05::0.8,(C:0.002,
30 #H1:0.002::0.2):0.048):0.01,A:0.06);
31 END;
32
33 BEGIN PHYLONET;
34 MCMC_SEQ -sgt (gt0,gt1) -snet net1 -sps 0.04;
35 END;

```

36 Gene trees are given in the Newick format, where the
 37 values after the colons are the branch lengths. Networks
 38 are given in the Rich Newick format which contains
 39 hybridization nodes denoted in “#H1”, “#H2”, ...,
 40 “#Hn”, where *n* is the number of reticulations in the
 41 network. The branch lengths, the population mutation
 42 rates, and the inheritance probabilities are specified after
 43 the first, second and third colons, respectively. Note that
 44 the units of branch lengths (for both trees and networks)
 45 can either be coalescent units, or the number of expected
 46 mutations per site, depending on the requirement of
 47 PhyloNet command. The population mutation rate is
 48 optional if the branch lengths are in the coalescent units,
 49 or a constant population mutation rate across all the
 50 branches is assumed. The inheritance probabilities are
 51 only relevant for the hybridization nodes.

52 The **InferNetwork_MP** command returns species
 53 networks and the corresponding extra lineages. The
 54 **InferNetwork_ML** command and its relatives return
 55 species networks and the corresponding likelihood values.
 56 The total number of returned networks can be specified
 57 via *-n* option for both commands. As we stated above,
 58 the user can either specify the maximum number
 59 of reticulation events or utilize the cross-validation,

60 bootstrap, information criteria to determine the model
 61 complexity when using maximum likelihood approach.

62 For Bayesian inference, the program outputs the log
 63 posterior probability, likelihood and prior for every
 64 sample. When the MCMC chain ends, the overall
 65 acceptance rates of the RJMCMC proposals and the
 66 95% credible set of species networks (the smallest set
 67 of all topologies that accounts for 95% of the posterior
 68 probability) are reported. For every topology in the
 69 95% credible set, the proportion of the topology being
 70 sampled, the maximum posterior value (MAP) and
 71 the corresponding MAP topology, the average posterior
 72 value and the averaged (branch lengths and inheritance
 73 probabilities) network are given. The model complexity
 74 is controlled mainly by the Poisson prior on the number
 75 of reticulations (see (Wen et al., 2016) for details). The
 76 Poisson distribution parameter can be tuned via “-pp”
 77 option.

CONCLUSION

78
 79 PhyloNet 3 is a comprehensive software package
 80 for phylogenetic network inference, particularly in the
 81 presence of incomplete lineage sorting. It implements
 82 maximum parsimony, maximum likelihood, and Bayesian
 83 inferences, in addition to a host of other features for
 84 analyzing phylogenetic networks and simulating data on
 85 them. The package is implemented in Java and is publicly
 86 available as an executable as well as source files.

87 In terms of the main aim of PhyloNet, which
 88 is the inference of phylogenetic networks, very few
 89 tools exist. TreeMix (Pickrell and Pritchard, 2012)
 90 is a very popular tool in the population genetics
 91 community. It uses allele frequency data and mainly
 92 targets analyses of admixtures among sub-populations
 93 of a single species. More recently, Bayesian inference
 94 of phylogenetic networks was implemented in BEAST2
 95 (Zhang et al., 2017) and inference of unrooted networks
 96 based on pseudo-likelihood was implemented in the
 97 PhyloNetworks software package (Solís-Lemus et al.,
 98 2017). However, in terms of implementing inference
 99 under different criteria and from different types of data,
 100 PhyloNet 3 is the most comprehensive.

101 As highlighted above, the challenge of computational
 102 tractability aside, the major challenge with network
 103 inference in general is determining the true number
 104 of reticulations and guarding against overfitting. In
 105 particular, phylogenetic networks are more complex
 106 models than trees and can always fit the data at least
 107 as well as trees do. Does this mean networks are simply
 108 over-parameterized models that should be abandoned in
 109 favor of trees? We argue that the answer is no. First,
 110 if the evolutionary history is reticulate, a tree-based
 111 method is unlikely to uncover the reticulation events
 112 (their number and locations). Second, even if one is
 113 interested in the species tree “despite reticulation,” a
 114 species tree inference method might not correctly recover
 115 the species tree (Solís-Lemus et al., 2016). Third, even
 116 when the true evolutionary history is strictly treelike,

1 the network structure could be viewed as a graphical
 2 representation of the variance around the tree structure.
 3 Insisting on a sparse network for convenience or ease of
 4 visual inspection is akin to insisting on a well-supported
 5 model no matter what the data says. Needless to say,
 6 one is interested in the true graph structure and not one
 7 that is more complicated simply because it assigns higher
 8 likelihood to the data. From our experience, the Bayesian
 9 approaches handle this challenge very well.

10 While we continue to improve the features and user-
 11 friendliness of the software package, the main direction
 12 we are currently pursuing is achieving scalability of the
 13 various inference methods in PhyloNet 3 to larger data
 14 sets in terms of the numbers of taxa as well as loci.

15 AVAILABILITY

16 PhyloNet is publicly available for download from
 17 <https://bioinfoc.rice.edu/phyloNet>. It can be installed
 18 and executed on any system with the Java Platform
 19 (Version 7.0 or higher). The current release includes
 20 source code, tutorials, example scripts, list of commands
 21 and useful links.

22 FUNDING

23 Development of the various functionalities in PhyloNet
 24 has been generously supported by DOE (DE-FG02-
 25 06ER25734), NIH (R01LM009494), and NSF (CCF-
 26 0622037, DBI-1062463, CCF-1302179, CCF-1514177)
 27 grants and Sloan and Guggenheim fellowships to L.N.
 28 Author contributions: D.W., Y.Y., and J.Z. did almost
 29 all the implementation and testing of the inference
 30 features in PhyloNet 3; D.W., J.Z. and L.N. wrote the
 31 paper.

32 ACKNOWLEDGEMENTS

33 We would like to thank the PhyloNet users for
 34 submitting bug reports, feature requests, and other
 35 comments that helped us significantly improve the
 36 software.

37 REFERENCES

38 Attias, H. 1999. Inferring parameters and structure of latent
 39 variable models by variational Bayes. Pages 21–30 *in*
 40 Proceedings of the Fifteenth conference on Uncertainty in
 41 artificial intelligence Morgan Kaufmann Publishers Inc.
 42 Bryant, D., R. Bouckaert, J. Felsenstein, N. A. Rosenberg, and
 43 A. RoyChoudhury. 2012. Inferring species trees directly from
 44 biallelic genetic markers: bypassing gene trees in a full coalescent
 45 analysis. *Molecular biology and evolution* 29:1917–1932.
 46 Degnan, J. H. and N. A. Rosenberg. 2009. Gene tree discordance,
 47 phylogenetic inference and the multispecies coalescent. *Trends*
 48 *in ecology & evolution* 24:332–340.
 49 Heled, J. and A. J. Drummond. 2010. Bayesian inference of species
 50 trees from multilocus data. *Molecular biology and evolution*
 51 27:570–580.
 52 Hudson, R. 2002. Generating samples under a Wright-Fisher
 53 neutral model of genetic variation. *Bioinformatics* 18:337–338.

Huson, D. H. and C. Scornavacca. 2012. Dendroscope 3: an
 54 interactive tool for rooted phylogenetic trees and networks.
 55 *Systematic biology* 61:1061–1067.
 56 Liu, L. 2008. Best: Bayesian estimation of species trees under the
 57 coalescent model. *Bioinformatics* 24:2542–2543.
 58 Maddison, W. P. 1997. Gene trees in species trees. *Systematic*
 59 *biology* 46:523–536.
 60 Maddison, W. P. and L. L. Knowles. 2006. Inferring phylogeny
 61 despite incomplete lineage sorting. *Systematic biology* 55:21–30.
 62 Mallet, J., N. Besansky, and M. Hahn. 2016. How reticulated are
 63 species? *BioEssays* 38:140–149.
 64 Mossel, E. and S. Roch. 2010. Incomplete lineage sorting:
 65 consistent phylogeny estimation from multiple loci. *IEEE/ACM*
 66 *Transactions on Computational Biology and Bioinformatics*
 67 (TCBB) 7:166–171.
 68 Nakhleh, L. 2010. Evolutionary phylogenetic networks: models
 69 and issues. Pages 125–158 *in* The Problem Solving Handbook
 70 for Computational Biology and Bioinformatics (L. Heath and
 71 N. Ramakrishnan, eds.). Springer, New York.
 72 Pickrell, J. K. and J. K. Pritchard. 2012. Inference of population
 73 splits and mixtures from genome-wide allele frequency data.
 74 *PLoS Genetics* 8:e1002967.
 75 Solís-Lemus, C., P. Bastide, and C. Ané. 2017. PhyloNetworks:
 76 a package for phylogenetic networks. *Molecular biology and*
 77 *evolution* 34:3292–3298.
 78 Solís-Lemus, C., M. Yang, and C. Ané. 2016. Inconsistency of
 79 species tree methods under gene flow. *Systematic Biology*
 80 65:843–851.
 81 Than, C. and L. Nakhleh. 2009. Species tree inference by
 82 minimizing deep coalescences. *PLoS Computational Biology*
 83 5:e1000501.
 84 Than, C., D. Ruths, and L. Nakhleh. 2008. PhyloNet: a software
 85 package for analyzing and reconstructing reticulate evolutionary
 86 relationships. *BMC bioinformatics* 9:322.
 87 Than, C. V. and N. A. Rosenberg. 2011. Consistency properties of
 88 species tree inference by minimizing deep coalescences. *Journal*
 89 *of Computational Biology* 18:1–15.
 90 Wen, D. and L. Nakhleh. 2017. Co-estimating reticulate
 91 phylogenies and gene trees from multi-locus sequence data.
 92 *Systematic Biology* syx085.
 93 Wen, D., Y. Yu, and L. Nakhleh. 2016. Bayesian inference
 94 of reticulate phylogenies under the multispecies network
 95 coalescent. *PLoS Genetics* 12:e1006006.
 96 Yu, Y., R. M. Barnett, and L. Nakhleh. 2013a. Parsimonious
 97 inference of hybridization in the presence of incomplete lineage
 98 sorting. *Systematic biology* Page syt037.
 99 Yu, Y., J. H. Degnan, and L. Nakhleh. 2012. The probability
 100 of a gene tree topology within a phylogenetic network
 101 with applications to hybridization detection. *PLoS genetics*
 102 8:e1002660.
 103 Yu, Y., J. Dong, K. J. Liu, and L. Nakhleh. 2014. Maximum
 104 likelihood inference of reticulate evolutionary
 105 histories. *Proceedings of the National Academy of Sciences*
 106 111:16448–16453.
 107 Yu, Y., C. Jermaine, and L. Nakhleh. 2016. Exploring phylogenetic
 108 hypotheses via gibbs sampling on evolutionary networks. *BMC*
 109 *genomics* 17:784.
 110 Yu, Y. and L. Nakhleh. 2015a. A distance-based method for
 111 inferring phylogenetic networks in the presence of incomplete
 112 lineage sorting. Pages 378–389 *in* International Symposium on
 113 Bioinformatics Research and Applications Springer.
 114 Yu, Y. and L. Nakhleh. 2015b. A maximum pseudo-likelihood
 115 approach for phylogenetic networks. *BMC genomics* 16:S10.
 116 Yu, Y., N. Ristic, and L. Nakhleh. 2013b. Fast algorithms and
 117 heuristics for phylogenomics under ILS and hybridization. *BMC*
 118 *Bioinformatics* 14:S6.
 119 Yu, Y., C. Than, J. H. Degnan, and L. Nakhleh. 2011. Coalescent
 120 histories on phylogenetic networks and detection
 121 of hybridization despite incomplete lineage sorting. *Systematic*
 122 *Biology* 60:138–149.
 123 Zhang, C., H. A. Ogilvie, A. J. Drummond, and T. Stadler. 2017.
 124 Bayesian inference of species networks from multilocus sequence
 125 data. *bioRxiv* Page 124982.
 126

- 1 Zhu, J., D. Wen, Y. Yu, H. Meudt, and L. Nakhleh. 2017.
- 2 Bayesian inference of phylogenetic networks from bi-allelic
- 3 genetic markers. *PLoS Computational Biology* (in press).
- 4 Zhu, J., Y. Yu, and L. Nakhleh. 2016. In the light of
- 5 deep coalescence: revisiting trees within networks. *BMC*
- 6 *Bioinformatics* 17:271–282.