Modern Computational Statistics

Lecture 20: Applications in Computational Biology

Cheng Zhang

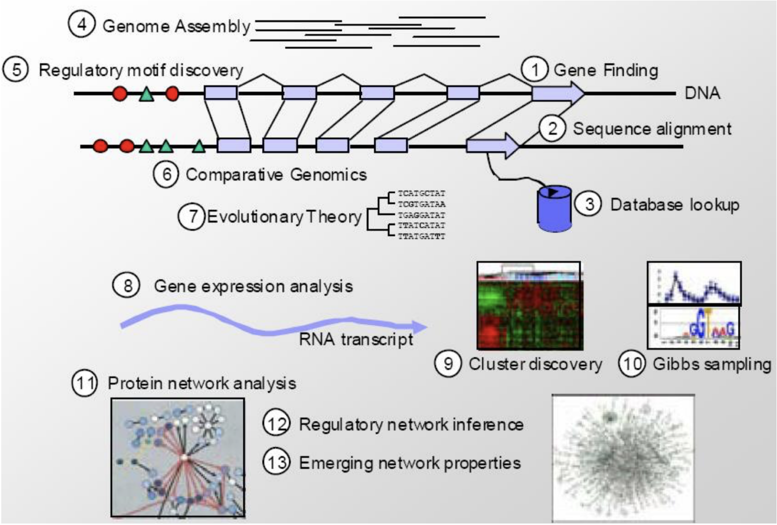
School of Mathematical Sciences, Peking University

December 09, 2019

Introduction

- While modern statistical approaches have been quite successful in many application areas, there are still challenging areas where the complex model structures make it difficult to apply those methods.
- In this lecture, we will discuss some of the recent advancement on statistical approaches for computational biology, with an emphasis on evolutionary models.

Challenges in Computational Biology



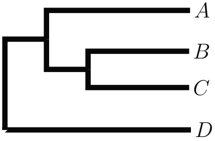
Adapted from Narges Razavian 2013

Phylogenetic Inference

The goal of **phylogenetic inference** is to reconstruct the evolution history (e.g., *phylogenetic trees*) from molecular sequence data (e.g., DNA, RNA or protein sequences)

Taxa	**Characters**
`Species A`	`ATGAACAT`
`Species B`	`ATGCACAC`
`Species C`	`ATGCATAT`
`Species D`	`ATGCATGC`

Molecular Sequence Data

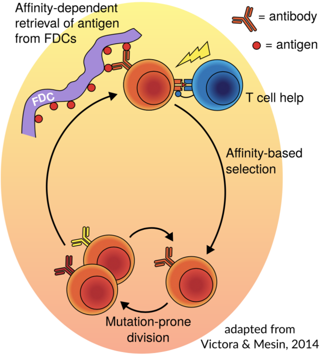


Phylogenetic Tree

Lots of modern biological and medical applications: *predict the evolution of influenza viruses and help vaccine design, etc.*

Example: B Cell Evolution

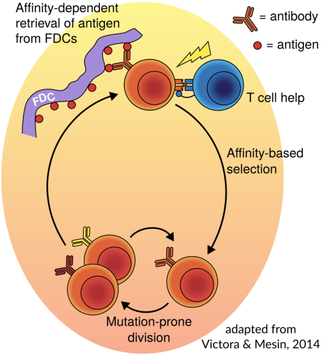
This happens inside of you!



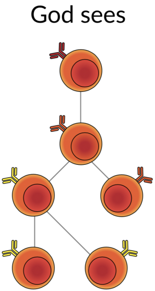
adapted from Victora & Mesin, 2014

Example: B Cell Evolution

This happens inside of you!

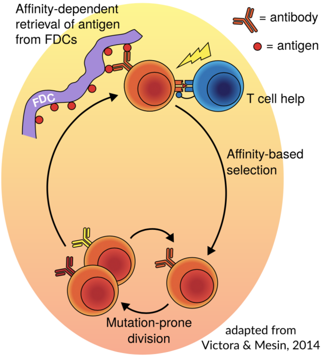


adapted from Victora & Mesin, 2014

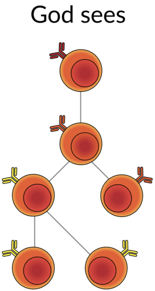


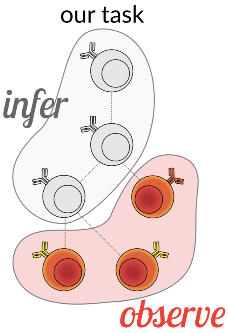
Example: B Cell Evolution

This happens inside of you!



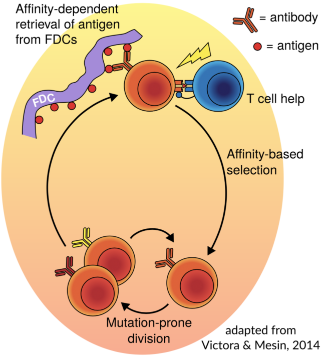
adapted from Victora & Mesin, 2014





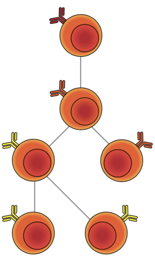
Example: B Cell Evolution

This happens inside of you!

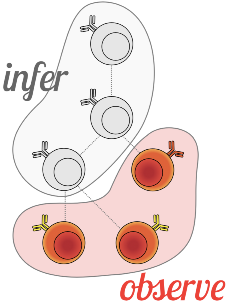


adapted from Victora & Mesin, 2014

God sees

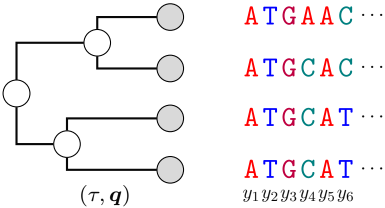


our task

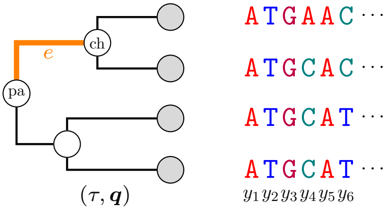


These inferences guide rational vaccine design.

Bayesian Phylogenetics



Bayesian Phylogenetics

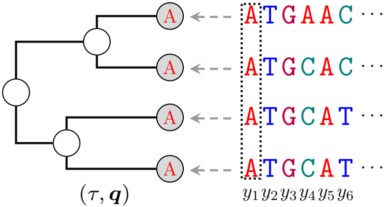


Evolution model:

$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Bayesian Phylogenetics



Evolution model:

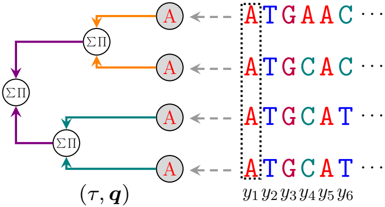
$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \qquad \eta(a_\rho^i) \prod_{(u,v)\in E(\tau)} P_{a_u^i a_v^i}(q_{uv})$$

Bayesian Phylogenetics



Evolution model:

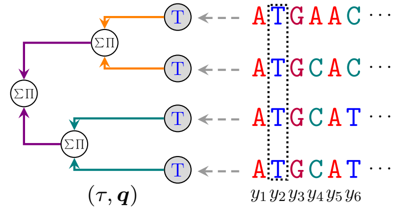
$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \sum_{a^i} \eta(a^i_\rho) \prod_{(u,v)\in E(\tau)} P_{a^i_u a^i_v}(q_{uv})$$

Bayesian Phylogenetics



Evolution model:

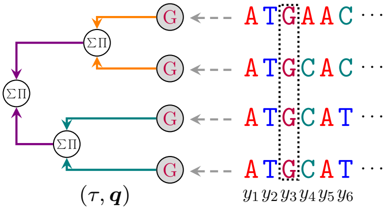
$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \prod_{i=1}^{M} \sum_{a^i} \eta(a_\rho^i) \prod_{(u,v)\in E(\tau)} P_{a_u^i a_v^i}(q_{uv})$$

Bayesian Phylogenetics



Evolution model:

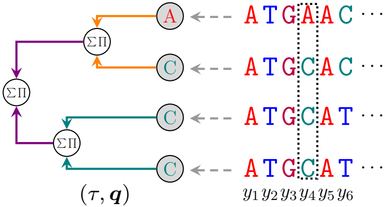
$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \prod_{i=1}^{M} \sum_{a^i} \eta(a_\rho^i) \prod_{(u,v)\in E(\tau)} P_{a_u^i a_v^i}(q_{uv})$$

Bayesian Phylogenetics



Evolution model:

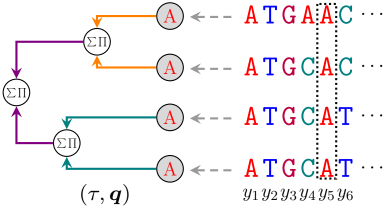
$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \prod_{i=1}^{M} \sum_{a^i} \eta(a_\rho^i) \prod_{(u,v)\in E(\tau)} P_{a_u^i a_v^i}(q_{uv})$$

Bayesian Phylogenetics



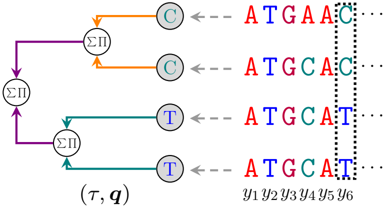
Evolution model:

$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \prod_{i=1}^{M} \sum_{a^i} \eta(a_\rho^i) \prod_{(u,v)\in E(\tau)} P_{a_u^i a_v^i}(q_{uv})$$



Evolution model:

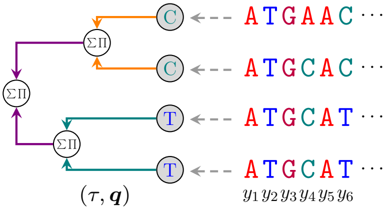
$$p(\text{ch}|\text{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \prod_{i=1}^{M} \sum_{a^i} \eta(a^i_\rho) \prod_{(u,v)\in E(\tau)} P_{a^i_u a^i_v}(q_{uv})$$

Bayesian Phylogenetics



Evolution model:

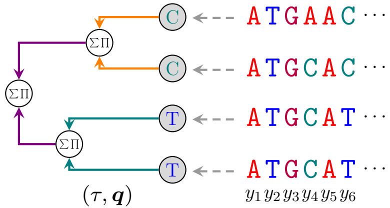
$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \prod_{i=1}^{M} \sum_{a^i} \eta(a_\rho^i) \prod_{(u,v)\in E(\tau)} P_{a_u^i a_v^i}(q_{uv})$$

Bayesian Phylogenetics



Evolution model:

$$p(\mathrm{ch}|\mathrm{pa}, q_e)$$

q_e: amount of evolution on e.

Likelihood

$$p(\boldsymbol{Y}|\tau, \boldsymbol{q}) = \prod_{i=1}^{M} \sum_{a^i} \eta(a_\rho^i) \prod_{(u,v)\in E(\tau)} P_{a_u^i a_v^i}(q_{uv})$$

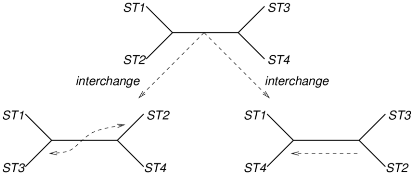
Given a proper prior distribution $p(\tau, \boldsymbol{q})$, the posterior is

$$p(\tau, \boldsymbol{q}|\boldsymbol{Y}) \propto p(\boldsymbol{Y}|\tau, \boldsymbol{q}) p(\tau, \boldsymbol{q}).$$

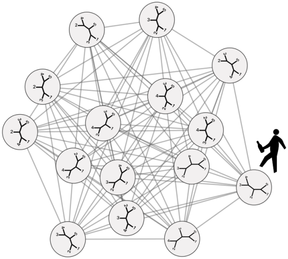
Markov chain Monte Carlo

Random-walk MCMC (MrBayes, BEAST):

- simple random perturbation (e.g., Nearest Neighborhood Interchange) to generate new state.



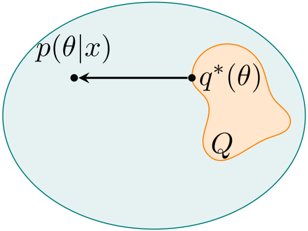
NNI



Challenges for MCMC

- Large search space: $(2n-5)!!$ unrooted trees (n taxa)
- Intertwined parameter space, low acceptance rate, hard to scale to data sets with many sequences.

Variational Inference



$$q^*(\theta) = \arg\min_{q \in Q} \mathrm{KL}\left(q(\theta) \| p(\theta|x)\right)$$

- VI turns inference into optimization
- Specify a variational family of distributions over the model parameters

$$Q = \{q_\phi(\theta); \phi \in \Phi\}$$

- Fit the variational parameters ϕ to minimize the distance (often in terms of KL divergence) to the exact posterior

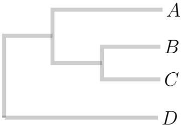
Evidence Lower Bound

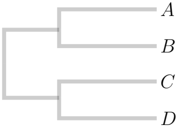
$$L(\theta) = \mathbb{E}_{q(\theta)}(\log p(x, \theta)) - \mathbb{E}_{q(\theta)}(\log q(\theta)) \leq \log p(x)$$

- KL is intractable; maximizing the **evidence lower bound** (ELBO) instead, which only requires the joint probability $p(x, \theta)$.
  - The ELBO is a lower bound on $\log p(x)$.
  - Maximizing the ELBO is equivalent to minimizing the KL.
- The ELBO strikes a balance between two terms
  - The first term encourages q to focus probability mass where the model puts high probability.
  - The second term encourages q to be diffuse.
- As an optimization approach, VI tends to be faster than MCMC, and is easier to scale to large data sets (via stochastic gradient ascent)

Subsplit Bayesian Networks

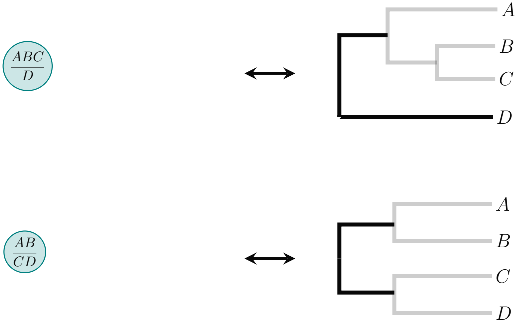
Inspired by previous works (Höhna and Drummond 2012, Larget 2013), we can decompose trees into local structures and encode the tree topology space via Bayesian networks!





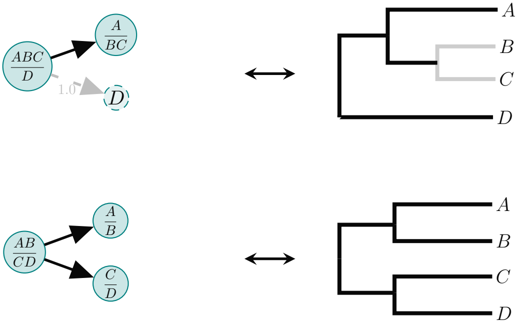
Subsplit Bayesian Networks

Inspired by previous works (Höhna and Drummond 2012, Larget 2013), we can decompose trees into local structures and encode the tree topology space via Bayesian networks!



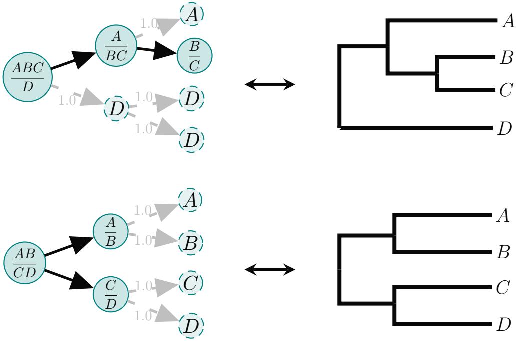
Subsplit Bayesian Networks

Inspired by previous works (Höhna and Drummond 2012, Larget 2013), we can decompose trees into local structures and encode the tree topology space via Bayesian networks!



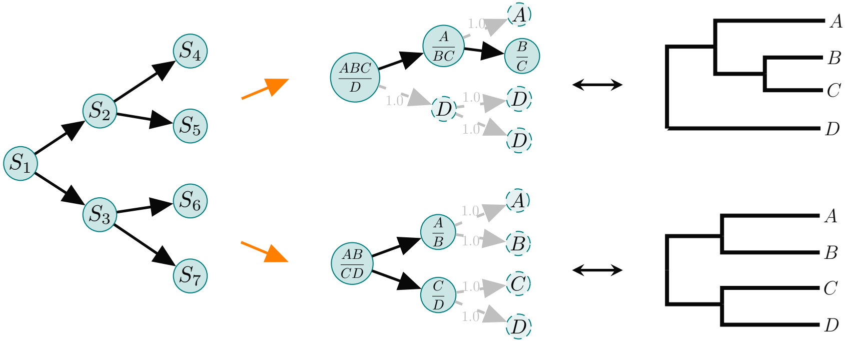
Subsplit Bayesian Networks

Inspired by previous works (Höhna and Drummond 2012, Larget 2013), we can decompose trees into local structures and encode the tree topology space via Bayesian networks!

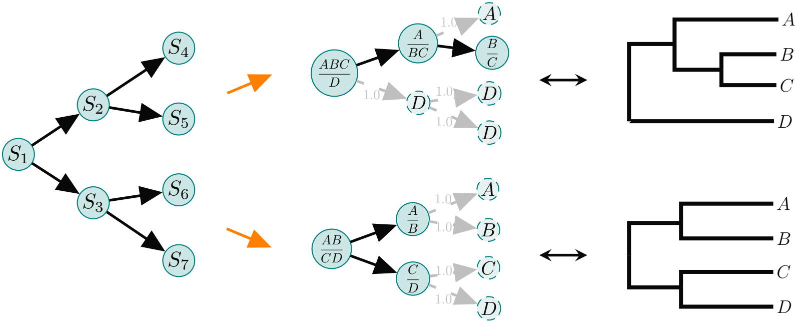


Subsplit Bayesian Networks

Inspired by previous works (Höhna and Drummond 2012, Larget 2013), we can decompose trees into local structures and encode the tree topology space via Bayesian networks!



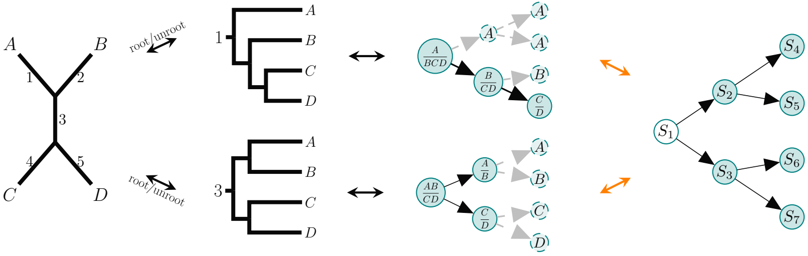
Probability Estimation Over Tree Topologies



Rooted Trees

$$p_{\mathrm{sbn}}(T=\tau)=p(S_1=s_1)\prod_{i>1}p(S_i=s_i|S_{\pi_i}=s_{\pi_i}).$$

Probability Estimation Over Tree Topologies



Unrooted Trees:

$$p_{\mathrm{sbn}}(T^{\mathrm{u}} = \tau) = \sum_{s_1 \sim \tau} p(S_1 = s_1) \prod_{i>1} p(S_i = s_i | S_{\pi_i} = s_{\pi_i}).$$

Tree Probability Estimation via SBNs

SBNs can be used to learn a probability distribution based on a collection of trees $T = \{T_1, \cdots, T_K\}$.

$$T_k = \{S_i = s_{i,k},\ i \geq 1\}, \quad k = 1, \ldots, K$$

Rooted Trees

- **Maximum Likelihood Estimates**: relative frequencies.

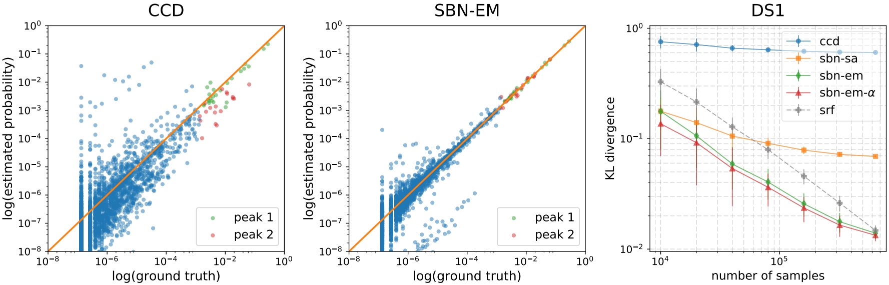
$$\hat{p}^{\mathrm{MLE}}(S_1 = s_1) = \frac{m_{s_1}}{K}, \quad \hat{p}^{\mathrm{MLE}}(S_i = s_i | S_{\pi_i} = t_i) = \frac{m_{s_i,t_i}}{\sum_{s \in \mathbb{C}_i} m_{s,t_i}}$$

Unrooted Trees

- **Expectation Maximization**

$$\hat{p}^{\mathrm{EM},(n+1)} = \arg\max_p \mathbb{E}_{p(S_1|T,\hat{p}^{\mathrm{EM},(\mathrm{n})})} \left(\log p(S_1) + \sum_{i>1} \log p(S_i|S_{\pi_i}) \right)$$

Example: Phylogenetic Posterior Estimation



[Zhang and Matsen, NeurIPS 2018]

- Compared to a previous method CCD (Larget, 2013), SBNs significantly reduce the biases for both high probability and low probability trees.
- SBNs perform better in the weak data regime.

Example: Phylogenetic Posterior Estimation

Data set	(#Taxa, #Sites)	Tree space size	Sampled trees	KL divergence to ground truth				
				SRF	CCD	**SBN-SA**	**SBN-EM**	**SBN-EM-α**
DS1	(27, 1949)	5.84×10^{32}	1228	0.0155	0.6027	0.0687	0.0136	**0.0130**
DS2	(29, 2520)	1.58×10^{35}	7	**0.0122**	0.0218	0.0218	0.0199	0.0128
DS3	(36, 1812)	4.89×10^{47}	43	0.3539	0.2074	0.1152	0.1243	**0.0882**
DS4	(41, 1137)	1.01×10^{57}	828	0.5322	0.1952	0.1021	0.0763	**0.0637**
DS5	(50, 378)	2.84×10^{74}	33752	11.5746	1.3272	0.8952	0.8599	**0.8218**
DS6	(50, 1133)	2.84×10^{74}	35407	10.0159	0.4526	**0.2613**	0.3016	0.2786
DS7	(59, 1824)	4.36×10^{92}	1125	1.2765	0.3292	0.2341	0.0483	**0.0399**
DS8	(64, 1008)	1.04×10^{103}	3067	2.1653	0.4149	0.2212	0.1415	**0.1236**

[Zhang and Matsen, NeurIPS 2018]

Remark: Unlike previous methods, SBNs are flexible enough to provide accurate approximations to real data posteriors!

Variational Bayesian Phylogenetic Inference

- Approximating Distribution:

$$\underbrace{Q_{\phi}(\tau)}^{\text{tree topology}}$$

Variational Bayesian Phylogenetic Inference

- Approximating Distribution:

$$Q_{\boldsymbol{\phi},\boldsymbol{\psi}}(\tau,\boldsymbol{q}) \triangleq \overset{\text{tree topology}}{\boxed{Q_{\boldsymbol{\phi}}(\tau)}} \cdot \overset{\text{branch length}}{\boxed{Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau)}}$$

Variational Bayesian Phylogenetic Inference

- Approximating Distribution:

$$Q_{\boldsymbol{\phi},\boldsymbol{\psi}}(\tau,\boldsymbol{q}) \triangleq \overset{\text{tree topology}}{\boxed{Q_{\boldsymbol{\phi}}(\tau)}} \cdot \overset{\text{branch length}}{\boxed{Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau)}}$$

- Multi-sample Lower Bound:

$$L^K(\boldsymbol{\phi},\boldsymbol{\psi}) = \mathbb{E}_{Q_{\boldsymbol{\phi},\boldsymbol{\psi}}(\tau^{1:K},\boldsymbol{q}^{1:K})} \log\left(\frac{1}{K}\sum_{i=1}^{K}\frac{p(\boldsymbol{Y}|\tau^i,\boldsymbol{q}^i)p(\tau^i,\boldsymbol{q}^i)}{Q_{\boldsymbol{\phi}}(\tau^i)Q_{\boldsymbol{\psi}}(\boldsymbol{q}^i|\tau^i)}\right)$$

Variational Bayesian Phylogenetic Inference

- Approximating Distribution:

$$Q_{\boldsymbol{\phi},\boldsymbol{\psi}}(\tau,\boldsymbol{q}) \triangleq \overset{\text{tree topology}}{\boxed{Q_{\boldsymbol{\phi}}(\tau)}} \cdot \overset{\text{branch length}}{\boxed{Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau)}}$$

- Multi-sample Lower Bound:

$$L^K(\boldsymbol{\phi},\boldsymbol{\psi}) = \mathbb{E}_{Q_{\boldsymbol{\phi},\boldsymbol{\psi}}(\tau^{1:K},\boldsymbol{q}^{1:K})} \log\left(\frac{1}{K}\sum_{i=1}^{K}\frac{p(\boldsymbol{Y}|\tau^i,\boldsymbol{q}^i)p(\tau^i,\boldsymbol{q}^i)}{Q_{\boldsymbol{\phi}}(\tau^i)Q_{\boldsymbol{\psi}}(\boldsymbol{q}^i|\tau^i)}\right)$$

- Use **stochastic gradient ascent** (SGA) to maximize the lower bound:

$$\hat{\boldsymbol{\phi}},\hat{\boldsymbol{\psi}} = \arg\max_{\boldsymbol{\phi},\boldsymbol{\psi}} \ L^K(\boldsymbol{\phi},\boldsymbol{\psi})$$

  - ϕ: VIMCO/RWS
  - ψ: The Reparameterization Trick

Structured Parameterization

SBNs Parameters

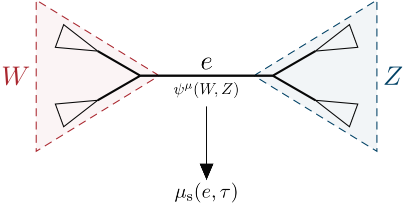
$$p(S_1 = s_1) = \frac{\exp(\phi_{s_1})}{\sum_{s_r \in \mathbb{S}_r} \exp(\phi_{s_r})}, \quad p(S_i = s | S_{\pi_i} = t) = \frac{\exp(\phi_{s|t})}{\sum_{s \in \mathbb{S}_{\cdot|t}} \exp(\phi_{s|t})}$$

Branch Length Parameters

$$Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau) = \prod_{e \in E(\tau)} p^{\text{Lognormal}}\left(q_e \mid \mu(e,\tau), \sigma(e,\tau)\right)$$

- *Simple Split*

  $\mu_{\mathrm{s}}(e,\tau) = \psi^{\mu}_{e/\tau},\ \sigma_{\mathrm{s}}(e,\tau) = \psi^{\sigma}_{e/\tau}.$



Structured Parameterization

SBNs Parameters

$$p(S_1 = s_1) = \frac{\exp(\phi_{s_1})}{\sum_{s_r \in \mathbb{S}_r} \exp(\phi_{s_r})}, \quad p(S_i = s | S_{\pi_i} = t) = \frac{\exp(\phi_{s|t})}{\sum_{s \in \mathbb{S}_{\cdot|t}} \exp(\phi_{s|t})}$$

Branch Length Parameters

$$Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau) = \prod_{e \in E(\tau)} p^{\text{Lognormal}}\left(q_e \mid \mu(e,\tau), \sigma(e,\tau)\right)$$

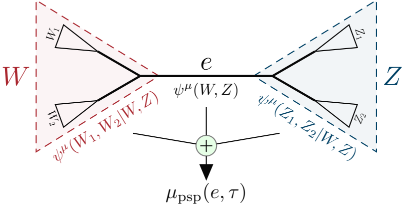
- *Simple Split*

$$\mu_{\mathrm{s}}(e,\tau) = \psi^{\mu}_{e/\tau}, \ \sigma_{\mathrm{s}}(e,\tau) = \psi^{\sigma}_{e/\tau}.$$

- *Primary Subsplit Pair* (PSP)

$$\mu_{\mathrm{psp}}(e,\tau) = \psi^{\mu}_{e/\tau} + \sum\nolimits_{s \in e/\!\!/\tau} \psi^{\mu}_{s}$$

$$\sigma_{\mathrm{psp}}(e,\tau) = \psi^{\sigma}_{e/\tau} + \sum\nolimits_{s \in e/\!\!/\tau} \psi^{\sigma}_{s}.$$


W
W_1
W_2
e
$\psi^{\mu}(W,Z)$
$\psi^{\mu}(W_1, W_2|W,Z)$
$\psi^{\mu}(Z_1, Z_2|W,Z)$
Z_1
Z_2
Z
$\mu_{\mathrm{psp}}(e,\tau)$

Stochastic Gradient Estimators

SBNs Parameters $\boldsymbol{\phi}$. With $\tau^j, \boldsymbol{q}^j \overset{\text{iid}}{\sim} Q_{\boldsymbol{\phi},\boldsymbol{\psi}}(\tau,\ \boldsymbol{q})$

- *VIMCO*. [Minh and Rezende, ICML 2016]

$$\nabla_{\boldsymbol{\phi}} L^K(\boldsymbol{\phi},\boldsymbol{\psi}) \simeq \sum_{j=1}^{K} \left(\hat{L}^K_{j|-j}(\boldsymbol{\phi},\boldsymbol{\psi}) - \tilde{w}^j\right) \nabla_{\boldsymbol{\phi}} \log Q_{\boldsymbol{\phi}}(\tau^j).$$

- *RWS*. [Bornschein and Bengio, ICLR 2015]

$$\nabla_{\boldsymbol{\phi}} L^K(\boldsymbol{\phi},\boldsymbol{\psi}) \simeq \sum_{j=1}^{K} \tilde{w}^j \nabla_{\boldsymbol{\phi}} \log Q_{\boldsymbol{\phi}}(\tau^j).$$

Branch Length Parameters $\boldsymbol{\psi}$. $g_{\boldsymbol{\psi}}(\boldsymbol{\epsilon}|\tau) = \exp(\boldsymbol{\mu}_{\boldsymbol{\psi},\tau} + \boldsymbol{\sigma}_{\boldsymbol{\psi},\tau} \odot \boldsymbol{\epsilon})$.

- *Reparameterization Trick*. Let $f_{\boldsymbol{\phi},\boldsymbol{\psi}}(\tau,\boldsymbol{q}) = \frac{p(\boldsymbol{Y}|\tau,\boldsymbol{q})p(\tau,\boldsymbol{q})}{Q_{\boldsymbol{\phi}}(\tau)Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau)}$.

$$\nabla_{\boldsymbol{\psi}} L^K(\boldsymbol{\phi},\boldsymbol{\psi}) \simeq \sum_{j=1}^{K} \tilde{w}^j \nabla_{\boldsymbol{\psi}} \log f_{\boldsymbol{\phi},\boldsymbol{\psi}}(\tau^j, g_{\boldsymbol{\psi}}(\boldsymbol{\epsilon}^j|\tau^j))$$

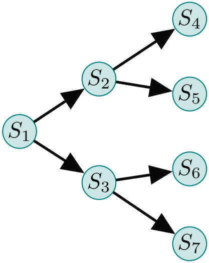
where $\tau^j \overset{\text{iid}}{\sim} Q_{\boldsymbol{\phi}}(\tau),\ \boldsymbol{\epsilon}^j \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{0},\boldsymbol{I})$.

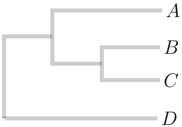
The VBPI Pipeline

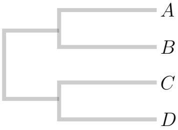
$Q_{\phi}(\tau)$

The VBPI Pipeline

Ancestral sampling for SBNs

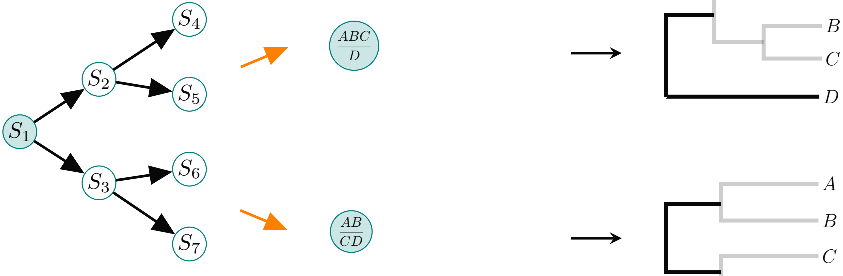






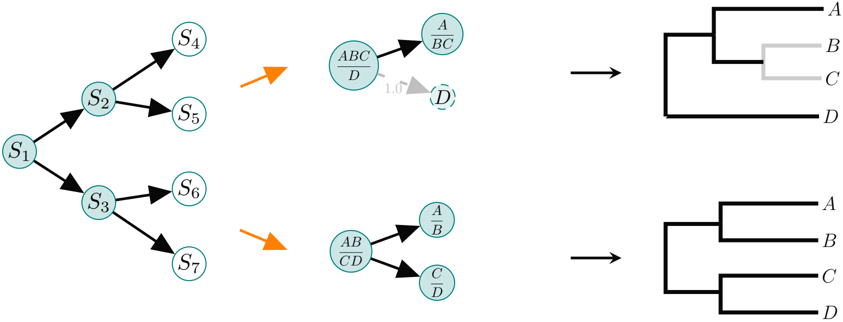
The VBPI Pipeline

Ancestral sampling for SBNs



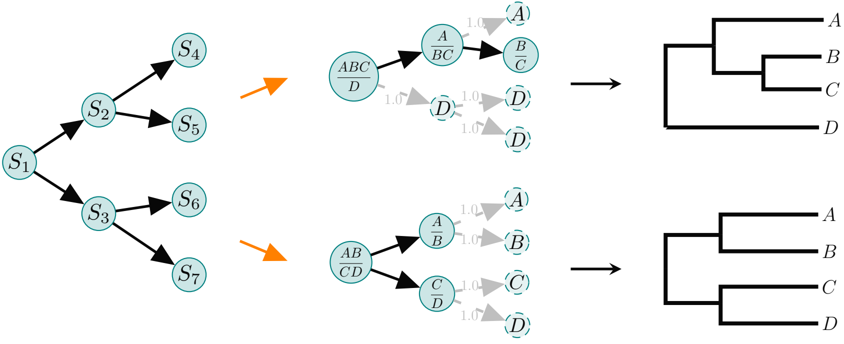
The VBPI Pipeline

Ancestral sampling for SBNs

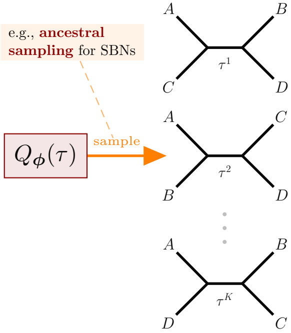


The VBPI Pipeline

Ancestral sampling for SBNs

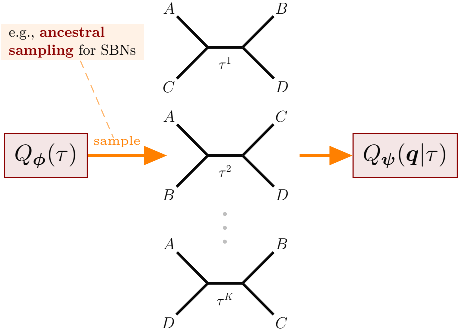


The VBPI Pipeline

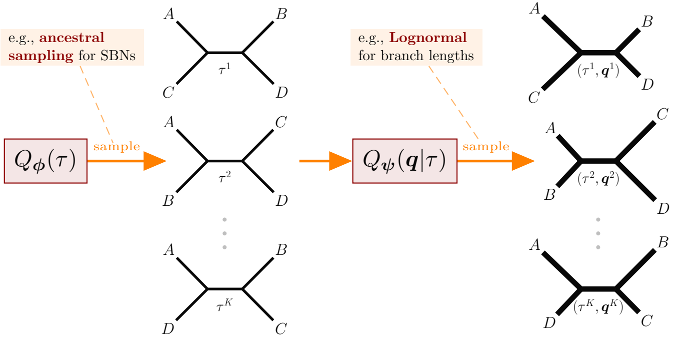


e.g., **ancestral sampling** for SBNs

The VBPI Pipeline


e.g., **ancestral sampling** for SBNs
$Q_{\boldsymbol{\phi}}(\tau)$
sample
A
B
C
D
τ^1
A
C
B
D
τ^2
A
B
D
C
τ^K
$Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau)$

The VBPI Pipeline



$Q_{\boldsymbol{\phi}}(\tau)$ sample

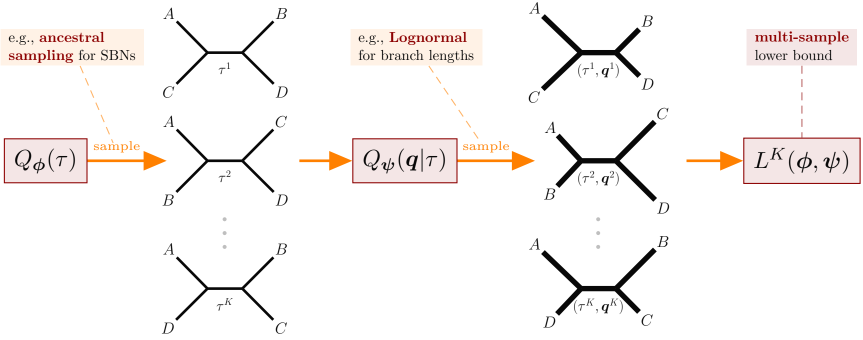
e.g., **ancestral sampling** for SBNs

$Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau)$ sample

e.g., **Lognormal** for branch lengths

A
B
C
D
τ^1
A
C
B
D
τ^2
A
B
D
C
τ^K
A
B
C
D
$(\tau^1, \boldsymbol{q}^1)$
A
C
B
D
$(\tau^2, \boldsymbol{q}^2)$
A
B
D
C
$(\tau^K, \boldsymbol{q}^K)$

The VBPI Pipeline



e.g., **ancestral sampling** for SBNs

$Q_{\boldsymbol{\phi}}(\tau)$

sample

e.g., **Lognormal** for branch lengths

$Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau)$

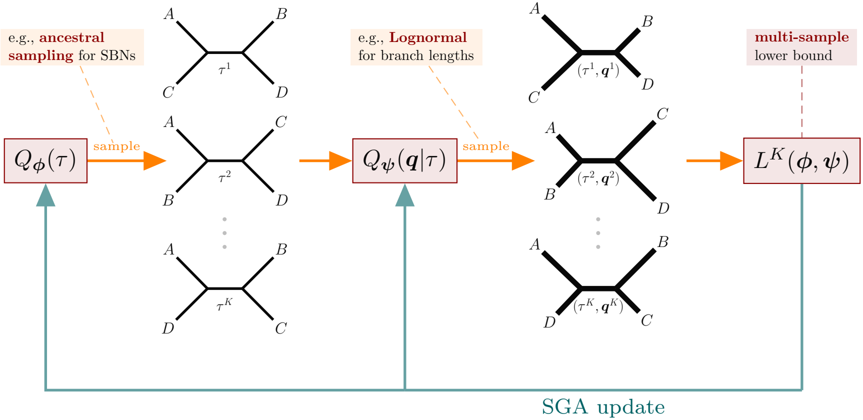
sample

multi-sample lower bound

$L^K(\boldsymbol{\phi}, \boldsymbol{\psi})$

A
B
C
D
τ^1
A
C
B
D
τ^2
A
B
D
C
τ^K
$(\tau^1, \boldsymbol{q}^1)$
$(\tau^2, \boldsymbol{q}^2)$
$(\tau^K, \boldsymbol{q}^K)$

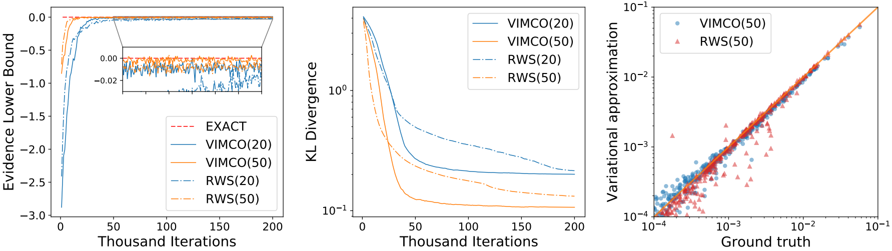
The VBPI Pipeline



e.g., **ancestral sampling** for SBNs
$Q_{\boldsymbol{\phi}}(\tau)$
sample
A
B
C
D
τ^1
τ^2
τ^K
e.g., **Lognormal** for branch lengths
$Q_{\boldsymbol{\psi}}(\boldsymbol{q}|\tau)$
sample
$(\tau^1, \boldsymbol{q}^1)$
$(\tau^2, \boldsymbol{q}^2)$
$(\tau^K, \boldsymbol{q}^K)$
multi-sample lower bound
$L^K(\boldsymbol{\phi}, \boldsymbol{\psi})$
SGA update

Performance on Synthetic Data

A simulated study on unrooted phylogenetic trees with 8 leaves (10395 trees). The target distribution is a random sample from the symmetric Dirichlet distribution $\mathrm{Dir}(\beta\mathbf{1})$, $\beta = 0.008$

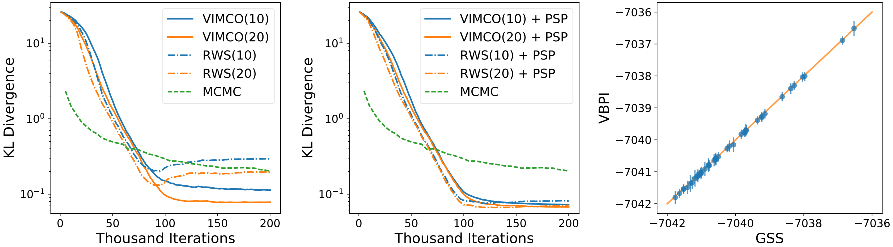


[Zhang and Matsen, ICLR 2019]

ELBOs approach 0 quickly ⇒ SBNs approximations are flexible.

More samples in the multi-sample ELBOs could be helpful.

Performance on Real Data



[Zhang and Matsen, ICLR 2019]

- More samples ⇒ better exploration ⇒ better approximation
- More flexible branch length distributions across tree topologies (PSP) ease training and improve approximation
- Outperform MCMC via much more efficient tree space exploration and branch length updates

Performance on Real Data

Data set	Marginal Likelihood (NATs)				
	VIMCO(10)	**VIMCO(20)**	**VIMCO(10)+PSP**	**VIMCO(20)+PSP**	SS
DS1	-7108.43(0.26)	-7108.35(0.21)	-7108.41(0.16)	**-7108.42(0.10)**	-7108.42(0.18)
DS2	-26367.70(0.12)	-26367.71(0.09)	**-26367.72(0.08)**	-26367.70(0.10)	-26367.57(0.48)
DS3	-33735.08(0.11)	-33735.11(0.11)	**-33735.10(0.09)**	-33735.07(0.11)	-33735.44(0.50)
DS4	-13329.90(0.31)	-13329.98(0.20)	**-13329.94(0.18)**	-13329.93(0.22)	-13330.06(0.54)
DS5	-8214.36(0.67)	-8214.74(0.38)	-8214.61(0.38)	-8214.55(0.43)	**-8214.51(0.28)**
DS6	-6723.75(0.68)	-6723.71(0.65)	-6724.09(0.55)	**-6724.34(0.45)**	-6724.07(0.86)
DS7	-37332.03(0.43)	-37331.90(0.49)	-37331.90(0.32)	**-37332.03(0.23)**	-37332.76(2.42)
DS8	-8653.34(0.55)	-8651.54(0.80)	**-8650.63(0.42)**	-8650.55(0.46)	-8649.88(1.75)

[Zhang and Matsen, ICLR 2019]

- Competitive to state-of-the-art (stepping-stone), dramatically reducing cost at test time: VBPI(1000) vs SS(100,000)
- PSP alleviates the demand for large samples, reducing computation while maintaining approximation accuracy

Conclusion

- We introduced **VBPI**, a general variational framework for Bayesian phylogenetic inference.
- **VBPI** allows efficient learning on both tree topology and branch lengths, providing competitive performance to MCMC while requiring much less computation.
- Can be used for further statistical analysis (e.g., marginal likelihood estimation) via importance sampling.
- There are many extensions, including more flexible branch length distributions, more general models, designing adaptive transition kernels in MCMC approaches, etc.

References

- Sebastian Hóhna and Alexei J. Drummond. Guided tree topology proposals for Bayesian phyloge- netic inference. Syst. Biol., 61(1):1–11, January 2012.
- Bret Larget. The estimation of tree posterior probabilities using conditional clade probability distributions. Syst. Biol., 62(4):501–511, July 2013.
- Zhang, C. and Matsen F. A., Generalizing Tree Probability Estimation via Bayesian Networks. In Advances in Neural Information Processing Systems, 2018.
- Zhang, C. and Matsen F. A., Variational Bayesian Phylogenetic Inference. In Proceedings of the 7th International Conference on Learning Representations, 2019.