



# MrBayes on GPU Cluster

Xiaoguang Liu

Nankai University, Tianjin, China

GTC ASIA 2011, Beijing, China

Nankai-Baidu  
Joint Laboratory



Parallel and Distributed  
Software Technology Lab



# Outline

- What's MrBayes
- Related Works
- Our contribution
- Experimental results
- Conclusion



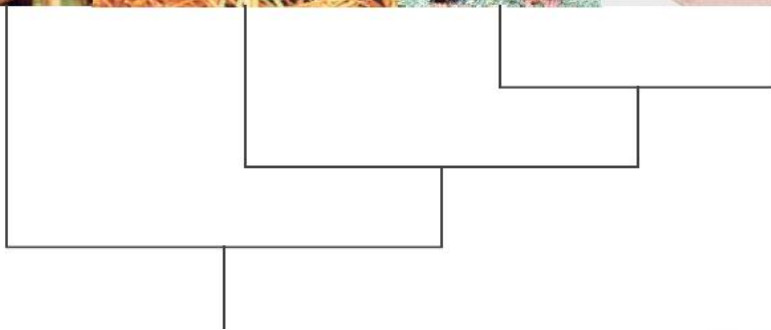
# Outline

- **What's MrBayes**
- Related Works
- Our contribution
- Experimental results
- Conclusion



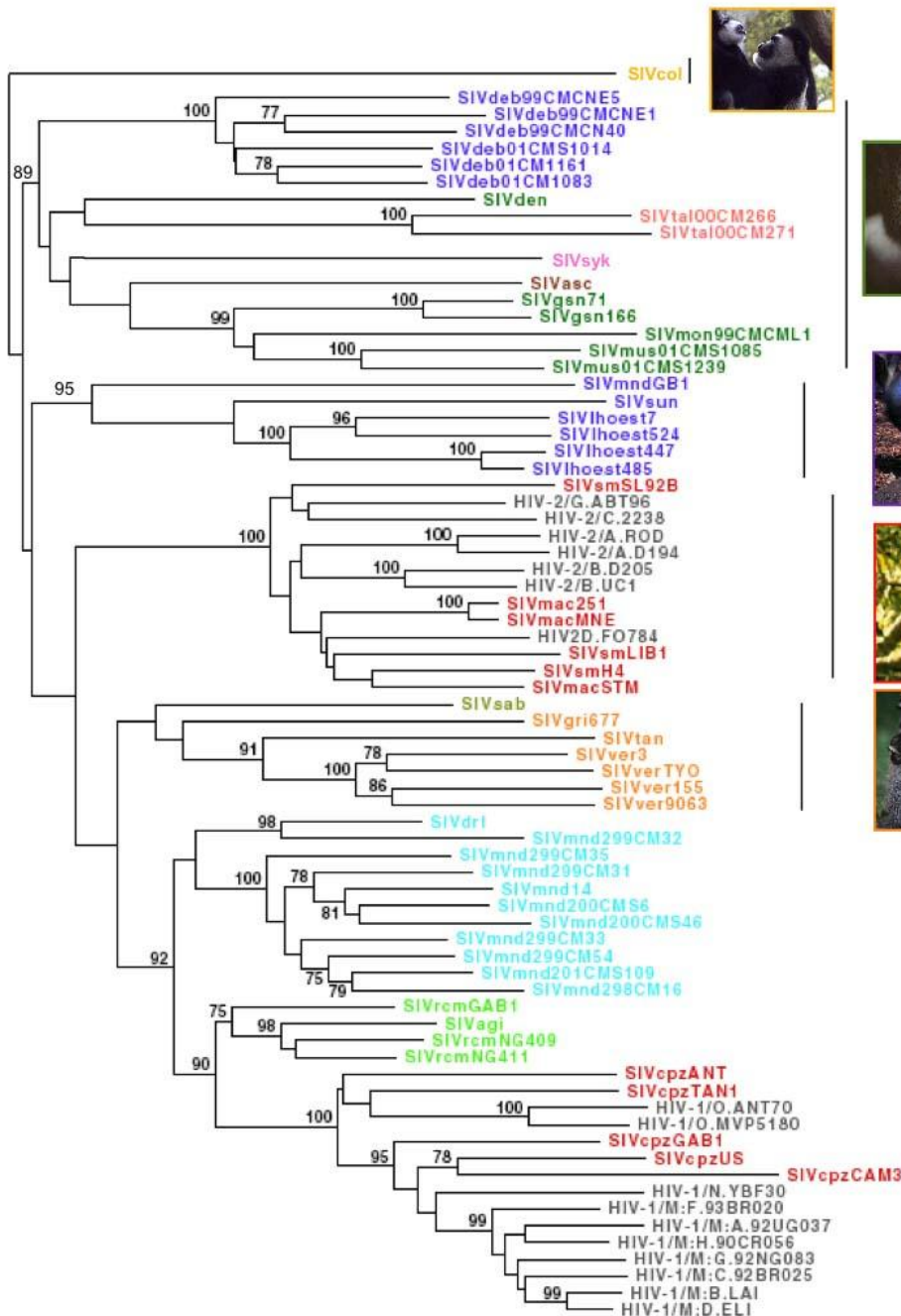
# MrBayes

- Given the DNA sequences of several organisms, MrBayes can be used to infer their phylogenetic ‘tree of life’.
- .More than 50% researchers use MrBayes to perform phylogenetic inference.



(posterior probability)  
 $\text{Pr}[\text{Tree}|\text{Data}]$





**Colobus**

**Cercopithecus**

**Mandrillus**

**Cercopithecus**

**Cercopithecus**

**Macaca\***

**Homo sapiens\*\***

**Chlorocebus**

**Mandrillus**

**M. sphinx**

**M. leucophaeus**

**C. torquatus**

**Pan**

**P.t. troglodytes**

**P.t. schweifurthii**

**Homo sapiens\*\***

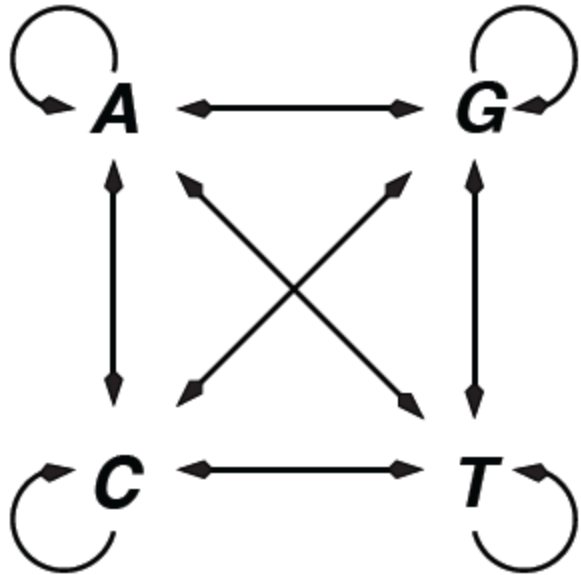
0.05 subst./site

# Metropolis coupled Markov chain Monte Carlo (MC<sup>3</sup>)

1. Initialization several Markov chains randomly
2. In each chain, :
  - ① Build a new tree via random perturb
  - ② Compute the accept probability
  - ③ Accept/Reject the new tree
3. Swap the data among chains at random



# Some concepts

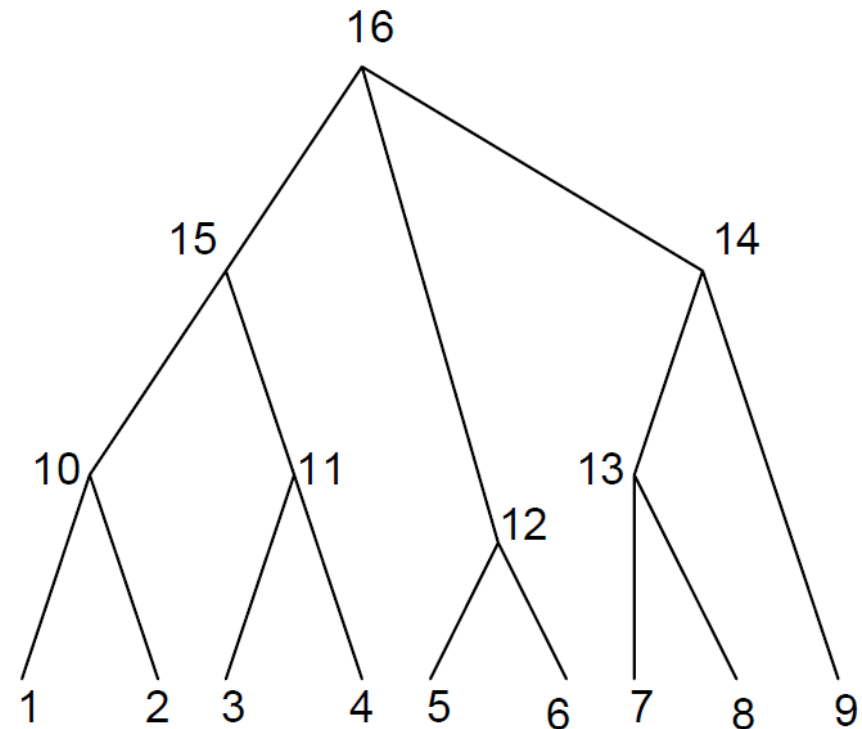
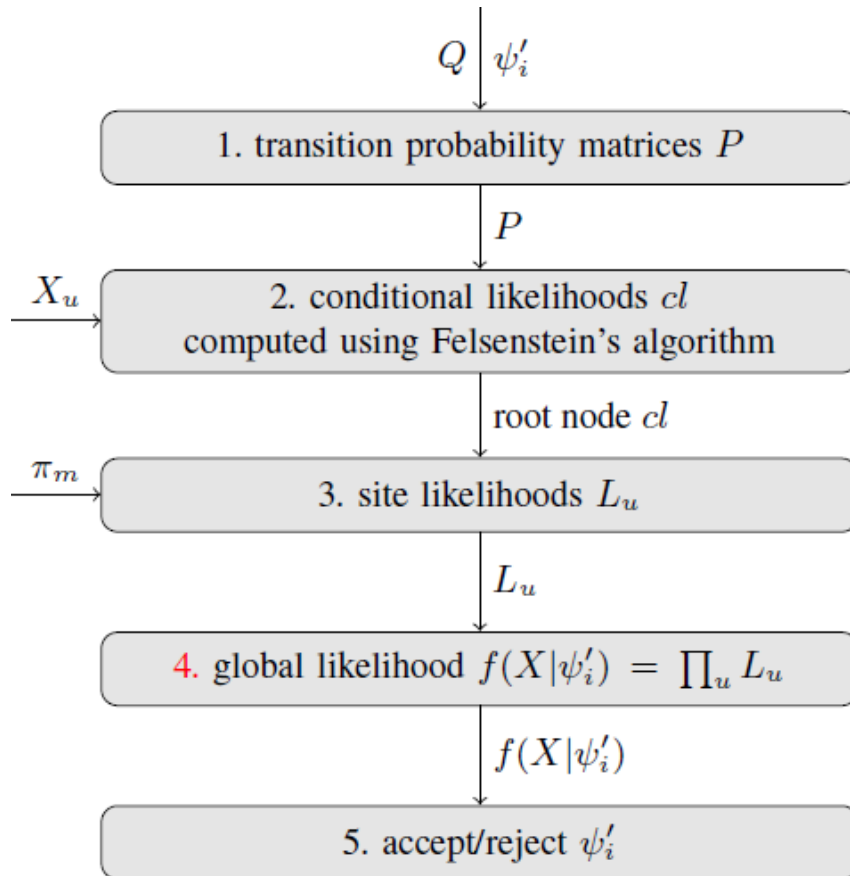


Q	A	C	G	T
A				
C				
G				
T				

- Q: instantaneous rate matrix
- transition probability matrix

$$P(t) = e^{Qt}$$

# Global likelihood



# Challenge

The screenshot shows the NCBI Taxonomy Browser interface. The browser address bar displays the URL <http://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi>. The page header includes the NCBI logo and navigation tabs for Entrez, PubMed, Nucleotide, Protein, Genome, Structure, PMC, Taxonomy, and Books. The search bar contains the text "as complete name" and has a "lock" checkbox checked. The "Display" dropdown is set to "3" levels using filter: "none". A list of search filters is provided, including Nucleotide, Nucleotide EST, Nucleotide GSS, Protein, Structure, Genome Sequences, Popset, SNP, Domains, GEO Datasets, GEO Expressions, UniGene, UniSTS, PubMed Central, Gene, HomoloGene, OMA, SRA Experiments, MapView, LinkOut, BLAST, TRACE, and Genome Projects. Below the search bar, the "Lineage (full): root; cellular organisms" is shown. The main content area displays a hierarchical list of organisms:

- o [Eukaryota](#) (eucaryotes) [71,235,521](#) [4,283](#) *Click on organism name to get more information.*
  - o [Alveolata](#) (alveolates) [1,093,774](#) [112](#)
    - o [Apicomplexa](#) (apicomplexans) [616,033](#) [96](#)
      - [Aconoidasida](#) [256,178](#) [81](#)
      - [Coccidia](#) [357,936](#) [15](#)
      - [Colpodellidae](#)
      - [Gregarina](#) [1,919](#)
      - [unclassified Apicomplexa](#)
      - [Apicomplexa incertae sedis](#)
      - [environmental samples](#)
    - o [Chromerida](#) [2,856](#) [2](#)
      - [Chromera](#) [2,856](#) [1](#)
      - [Chromerida sp. RM11](#) [1](#)
    - o [Ciliophora](#) (ciliates) [269,564](#) [11](#)
      - [Intramacronucleata](#) [269,564](#) [11](#)
      - [Postciliodesmatophora](#)
      - [unclassified Ciliophora](#)
      - [environmental samples](#)
    - o [Dinophyceae](#) (dinoflagellates) [173,496](#) [2](#)
      - [Blastodiales](#)
      - [Coccidiales](#)
      - [Dinophysiales](#)
      - [Gonyaulacales](#) [53,960](#)
      - [Gymnodinales](#) [85,914](#)
      - [Lophodinales](#)

# Outline

- What's MrBayes
- **Related Works**
- Our contribution
- Experimental results
- Conclusion



# Parallel algorithms on CPU

- pMC<sup>3</sup> Altekar *et al.* (2004),
  - Assign Markov chains to all CPU cores as even as possible
- hMC<sup>3</sup> Feng *et al.* (2006), Zhou *et al.* (2010),
  - Based on pMC<sup>3</sup>, a Markov chain also can be divided into several parts, it make the jobs afforded by CPU cores are more even.



# Parallel algorithms on GPU

- gMC<sup>3</sup> Pratas *et al.*, 2009
  - Every Markov Chain is distributed to one or more GPU thread to use the power of GPU
- Contribution : MrBayes can benefit from GPU
- Problem : poor performance

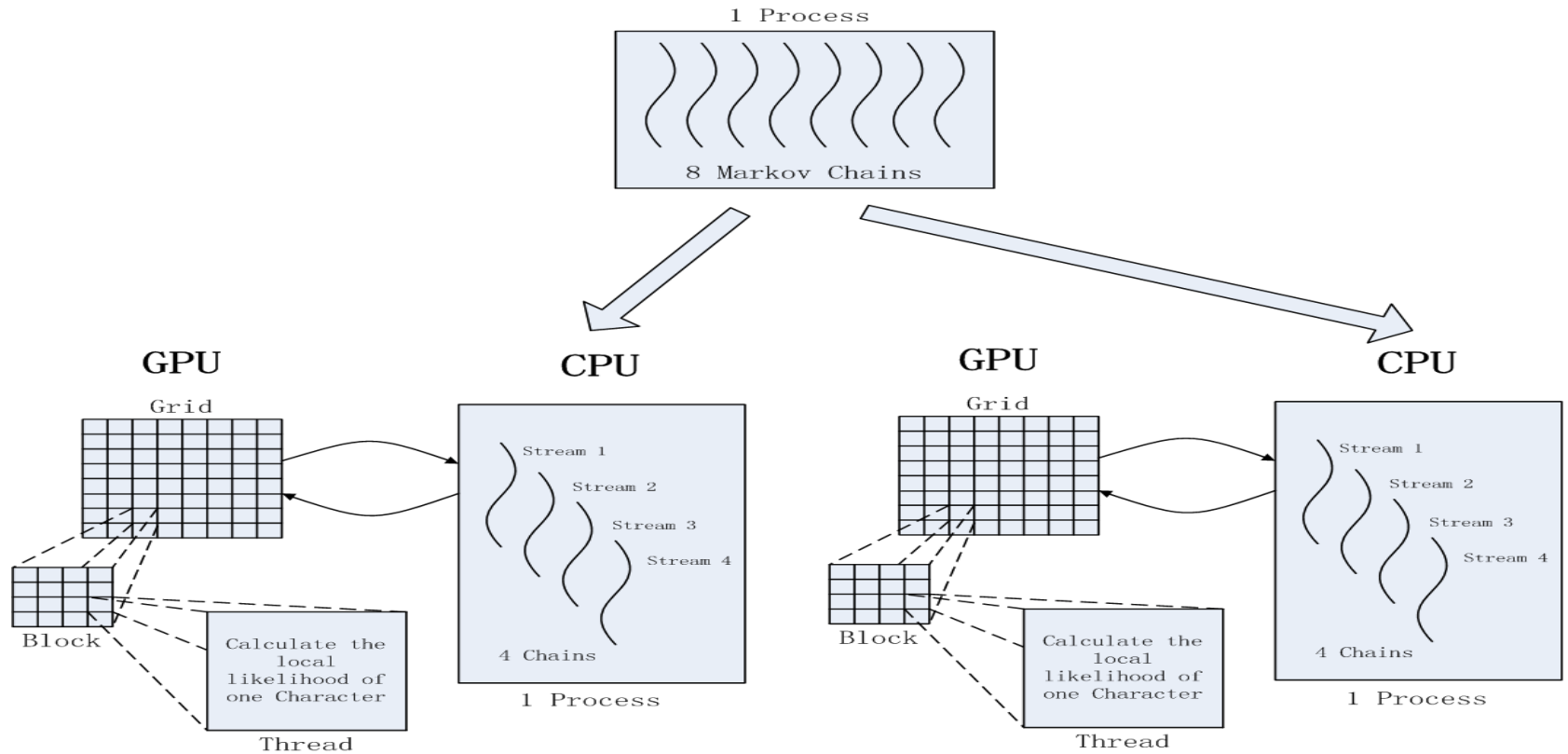


# Outline

- What's MrBayes
- Related Works
- **Our contribution**
- Experimental results
- Conclusion

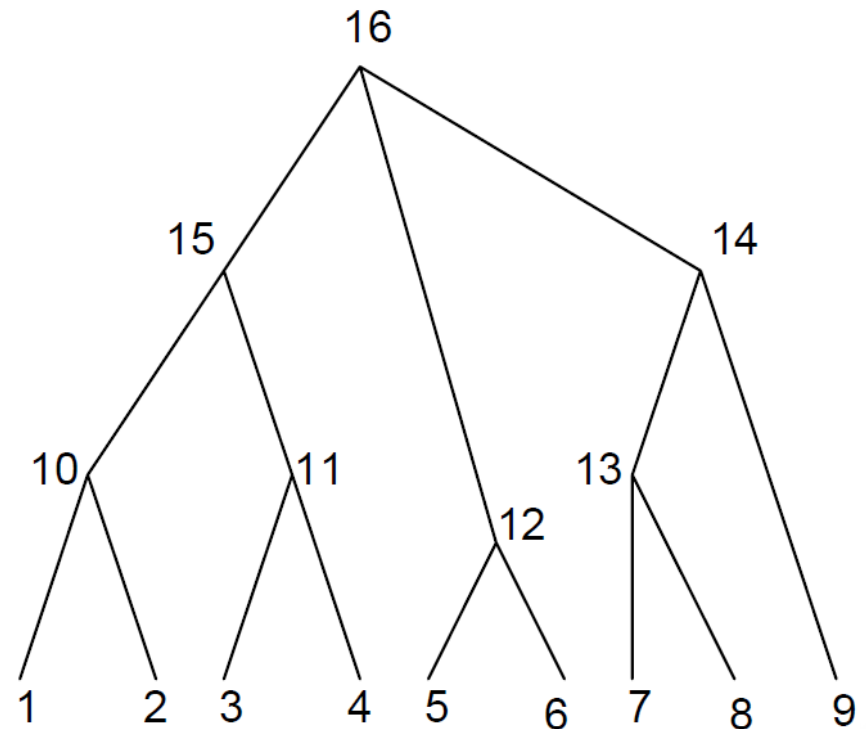
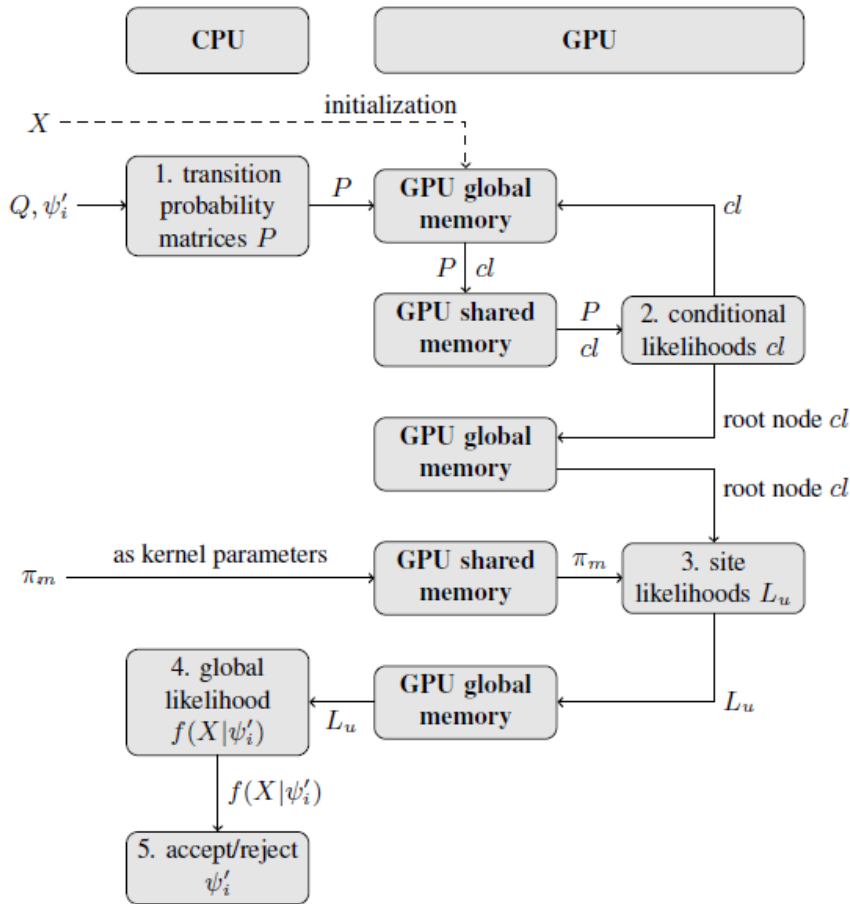


# CPU/GPU cooperate parallelism



- GPU and CPU co-processing multigrain parallel algorithm.

# Our algorithm: nMC<sup>3</sup>



# Outline

- What's MrBayes
- Related Works
- Our contribution
- **Experimental results**
- Conclusion



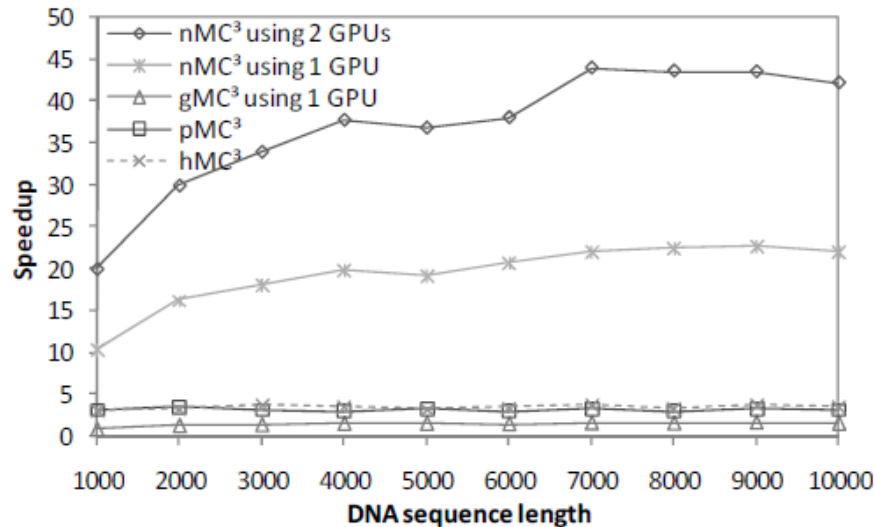
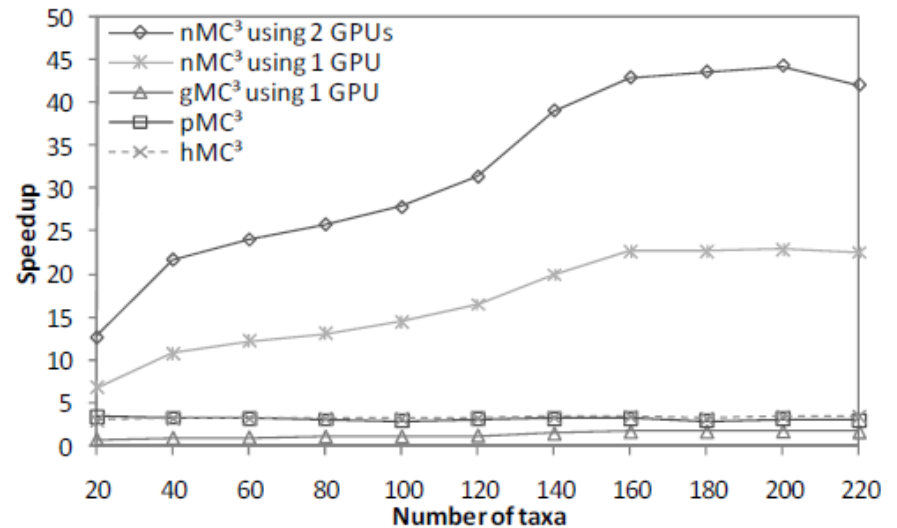
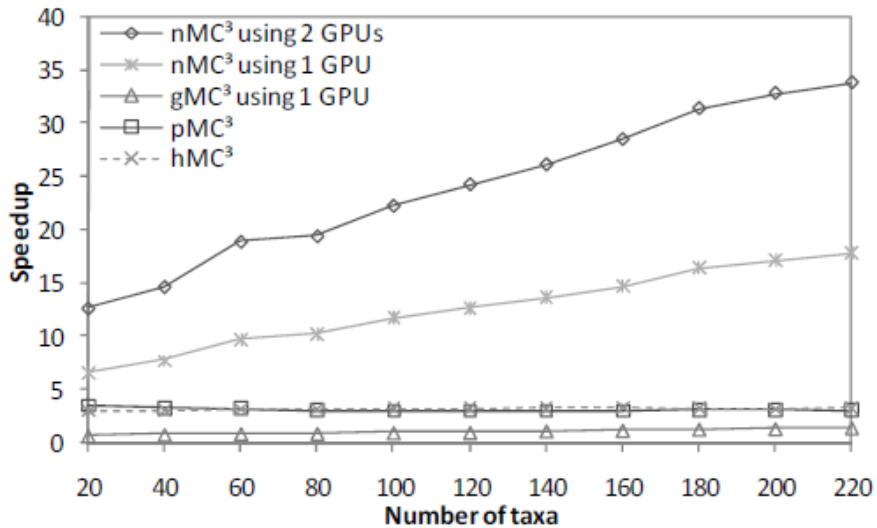
# Speedup

Dataset	# Taxa	DNA length	# Generations	Run-time (sec.)					
				MrBayes MC <sup>3</sup>	hMC <sup>3</sup> /pMC <sup>3</sup>	gMC <sup>3</sup> with 1 GPU	nMC <sup>3</sup> with 1 GPU	nMC <sup>3</sup> with 2 GPUs	
1	26	1546	1000000	4238	1227	10034	1343	731	
2	37	2238	1000000	14962	4363	20654	1784	943	
3	111	1506	500000	17234	5352	20321	1765	929	
4	234	1790	100000	15160	4275	10015	797	414	
5	288	3386	100000	31700	10456	19151	–	750	

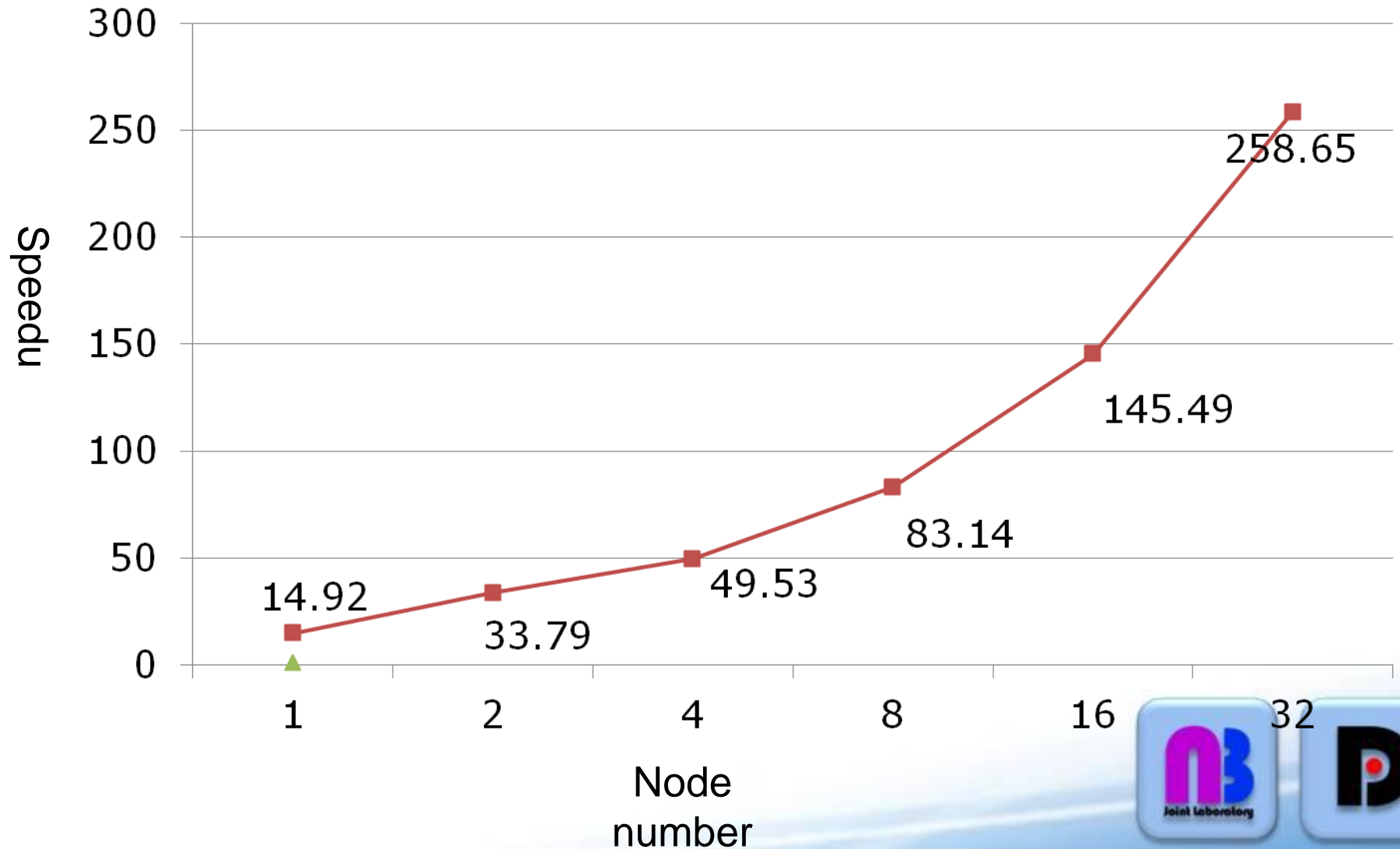
Dataset	Speedup		4-Speedup		g-Speedup	
	1 GPU	2 GPUs	1 GPU	2 GPUs	1 GPU	2 GPUs
1	3.2	5.8	0.9	1.7	7.5	13.7
2	8.4	15.9	2.4	4.6	11.6	21.9
3	9.8	18.6	3.0	5.8	11.5	21.9
4	19.0	36.6	5.4	10.3	12.6	24.2
5	–	42.3	–	13.9	–	25.5



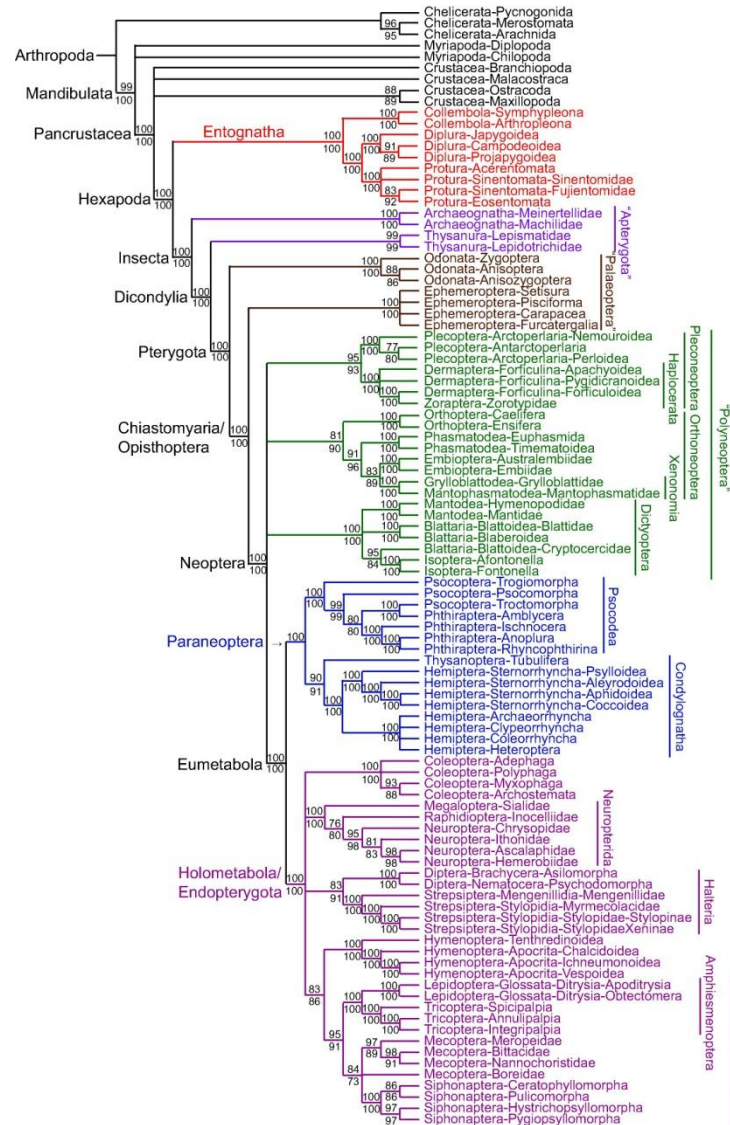
# Scalability



# Results on Tianhe 1-A



# Phylogenetic Tree

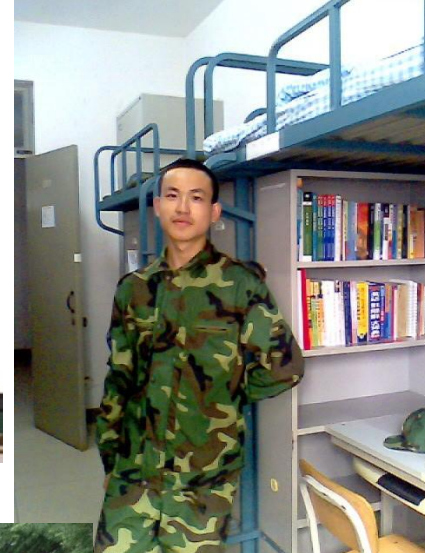


# Conclusion

- Give an efficient GPU-based MrBayes program
- It has good speedup
- It can decrease experimental time
- It can calculate larger dataset
- In the future ,run some large dataset on Tianhe,



Jianfu Zhou  
Douglas Stones  
Hongju Xia  
Jie Bao  
Qiang Xie



**ACKNOWLEDGMENT !**





**DOWNLOADS**

**527**

In the selected date range

**TOP COUNTRY**

**United States**

23% of downloaders

**TOP OS**

**Windows**

46% of downloaders

Country ↕	Downloads ▲
1. United States	122
2. China	111
3. Taiwan	39
4. Brazil	37
5. Germany	20
6. Australia	19
7. Spain	18
8. Canada	15
9. Russia	14
10. United Kingdom	13
11. India	12
12. Italy	11
13. France	11
14. Sweden	10
15. Switzerland	7
16. Japan	7