

Tam Birleşme (Coalescence) Modelli Filocoğrafya Uygulaması

İsmail K. Sağlam

Hacettepe Üniversitesi, Biyoloji Bölümü, Ekolojik Bilimler Araştırma Laboratuvarları

Cevaplamak İstedığımız Sorular

- ***Hopterix maximus* populasyonları arası farklı göç yollarının sınanması**
 - Migrate kullanarak Theta ve m parametrelerinin hesaplanması
 - Bayes Faktör kullanarak alternatif hipotezlerin (göç yollarının sınanması)
- ***Phonochorion* populasyonlarında ortak ataya varma zaminin hesaplanması**
 - Beast kullanarak standard birleşme modeli altında Tmrca hesaplaması
 - Beast kullanarak demografik büyüme birleşme modeli altında Tmrca hesaplaması
 - Beast kullanarak Bayesian Skyline Plot'ların çizilmesi ve Nef in zaman içindeki değişiminin modellenmesi

Hopterix maximus populasyonları arası farklı göç yollarının sınanması

- **Migrate:** <http://popgen.sc.fsu.edu/Migrate/Download.html>
- **Veri Seti:** *Hopterix maximus* 3 POP veri seti
- **Amaç:** Doğu Karadenizde yayılan *Hopterix maximus*'un Trabzon, Rize ve Artvin populasyonları arasındaki göç örüntülerini tanımlayan farklı hipotezlerin (modellerin) sınanması
 - 1. Model: Her populasyon her biri ile bağlantılı 3 Populasyon büyüklüğü parametresi (θ), 6 göç parametresi
 - 2. Model: Komşu populasyon modeli, simetrik göç, 3 Populasyon büyüklüğü parametresi (θ), 2 göç parametresi
 - 3. Model: Artvin izolasyon modeli, 3 populasyon büyüklüğü parametresi, 2 göç parametresi

	1 (Artv)	2 (Riz)	3 (Trbzn)
1 (Artvn)	Theta	$2 \rightarrow 1$	$3 \rightarrow 1$
2 (Riz)	$1 \rightarrow 2$	Theta	$3 \rightarrow 2$
3 (Trbzn)	$1 \rightarrow 3$	$2 \rightarrow 3$	Theta

Model 1: 3 theta 6 göç (m)

Migrate kodu: *** ** *

	1 (Artv)	2 (Riz)	3 (Trbzn)
1 (Artvn)	Theta	$2 \rightarrow 1$	NA
2 (Riz)	$1 \rightarrow 2$	Theta	$3 \rightarrow 2$
3 (Trbzn)	NA	$2 \rightarrow 3$	Theta

Model 2: 3 theta 2 göç (m)

Migrate kodu: *s0 s*s 0s*

$$1 \rightarrow 2 = 2 \rightarrow 1$$

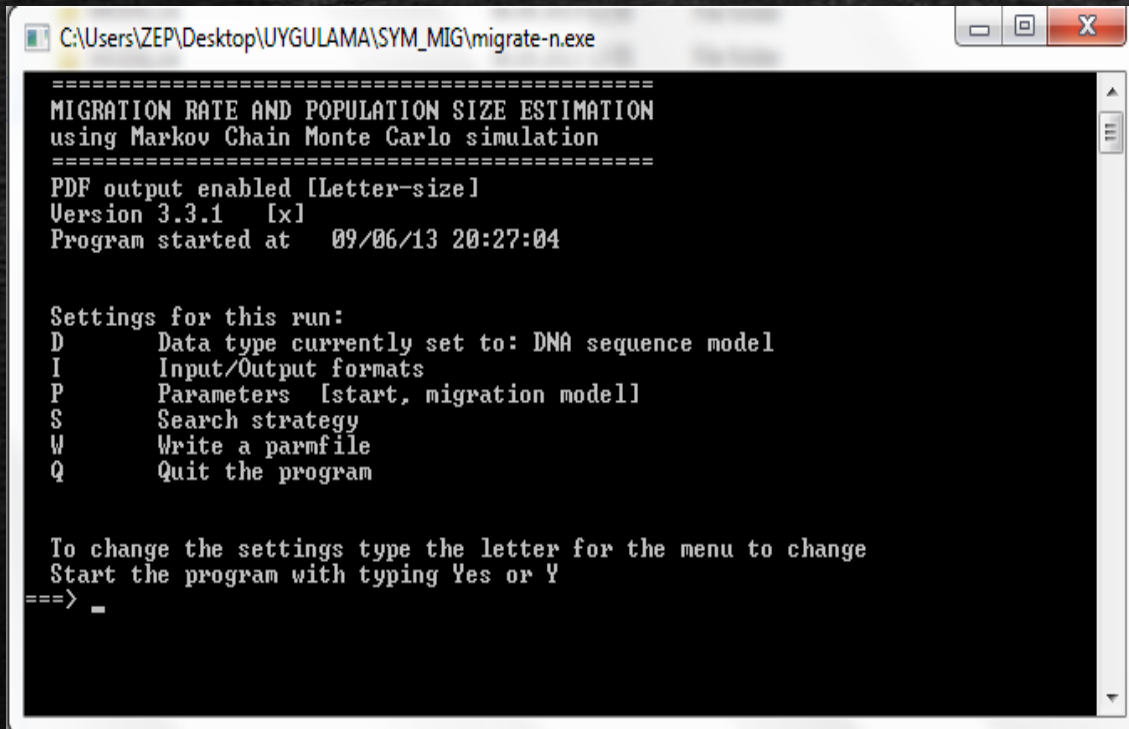
$$2 \rightarrow 3 = 3 \rightarrow 2$$

	1 (Artv)	2 (Riz)	3 (Trbzn)
1 (Artvn)	Theta	$2 \rightarrow 1$	NA
2 (Riz)	$1 \rightarrow 2$	Theta	NA
3 (Trbzn)	NA	NA	Theta

Model 3: 3 theta 2 göç (m)

Migrate kodu: **0 *00 00*

Migrate Programını Aç: Panik Yok!!!



```
C:\Users\ZEP\Desktop\UYGULAMA\SYM_MIG\migrate-n.exe

=====
MIGRATION RATE AND POPULATION SIZE ESTIMATION
using Markov Chain Monte Carlo simulation
=====
PDF output enabled [Letter-size]
Version 3.3.1 [x]
Program started at 09/06/13 20:27:04

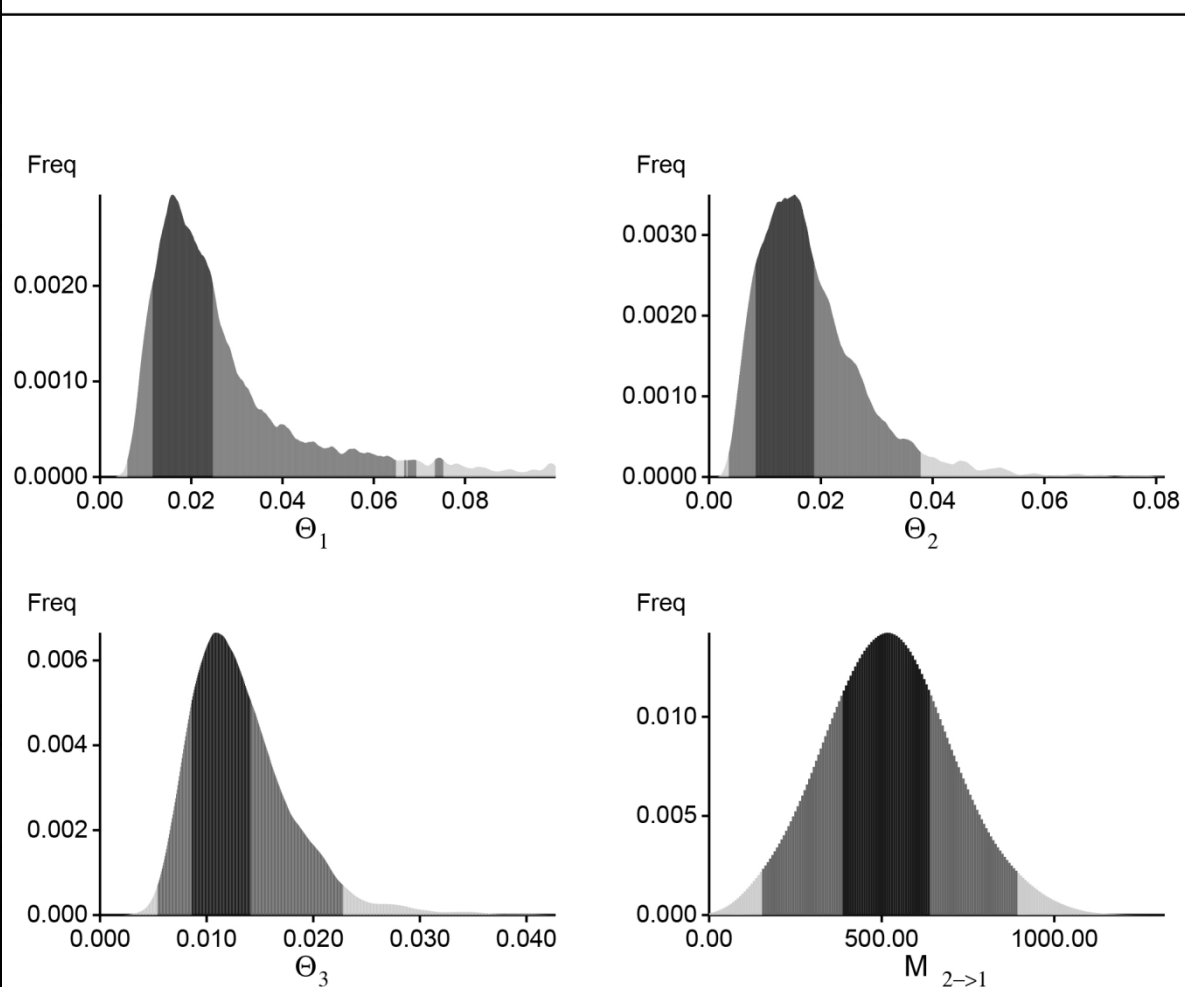
Settings for this run:
D Data type currently set to: DNA sequence model
I Input/Output formats
P Parameters [start, migration model]
S Search strategy
W Write a parmfile
Q Quit the program

To change the settings type the letter for the menu to change
Start the program with typing Yes or Y
==> -
```

- “i” tuşuna basarak input/output menüsüne girin ve datafile ismini hopterix3pop’a outputfile ismini model1’e çevirin. “Y” tuşu ile ana menüye dönün.
- P tuşu ile parameters menüsüne 8 ile göç modellerini tanıttın
- S tuşu ile search strategy menüsüne girin 7 seçeneği ile migration prior distribution’ı uniform 0 100000 10000 yapın. Heating seçeneğini açarak 1 1.5 3 5 sıcaklıklarında 4 zincir oluşturun
- W tuşuna basarak analiz dosyası yazın
- Y tuşuna basarak analize başlayın

Sonuç Dosyası

Bayesian Analysis: Posterior distribution over all loci



Ardı Olasılık Dağılımları: Bütün parametre tahminleri tek tepeli dağılım olmalı

Model altında verinin **Log-olasılığı** (sınır olabilesi)

Log-Probability of the data given the model (marginal likelihood)

Use this value for Bayes factor calculations:

$BF = \text{Exp}[\ln(\text{Prob}(D \mid \text{thisModel})) - \ln(\text{Prob}(D \mid \text{otherModel}))]$

or as $LBF = 2 (\ln(\text{Prob}(D \mid \text{thisModel})) - \ln(\text{Prob}(D \mid \text{otherModel})))$

shows the support for thisModel]

Method	$\ln(\text{Prob}(D \mid \text{Model}))$	Notes
Thermodynamic integration	(not estimated [no heating])	(1)
Harmonic mean	-1036.119803	(2)

Bayes Faktörleri Kullanarak Model Karşılaştırması

Bayes Faktör Formülü

$$LBF = 2(\ln mL(\text{model}_1) - \ln L(\text{model}_2))$$

-2'den daha küçük değerler "model2" ye yönelik tercih belirtirken, 2'den daha büyük değerler "model 1" e yönelik tercih belirtir. Modelleri log olabilirlik değerlerini kullanarak sıralayabiliriz. En düşük değere sahip model en iyi modeldir.

Modelin Olasılığı

$$\text{Prob}(\text{model}_i) = \frac{mL_{\text{model}_i}}{\sum_j^n mL_{\text{model}_j}}$$

Model olasılıklarınıda hesaplayabiliriz. Model olasılıkları her bir modelin Log olabilirliğini bütün sınanan modellerin toplam Log olabilirliğine bölerek hesaplarız:

Bayes Faktör ve Log olabilirlik değerlerine altında örnek model karşılaştırma çizelgesi

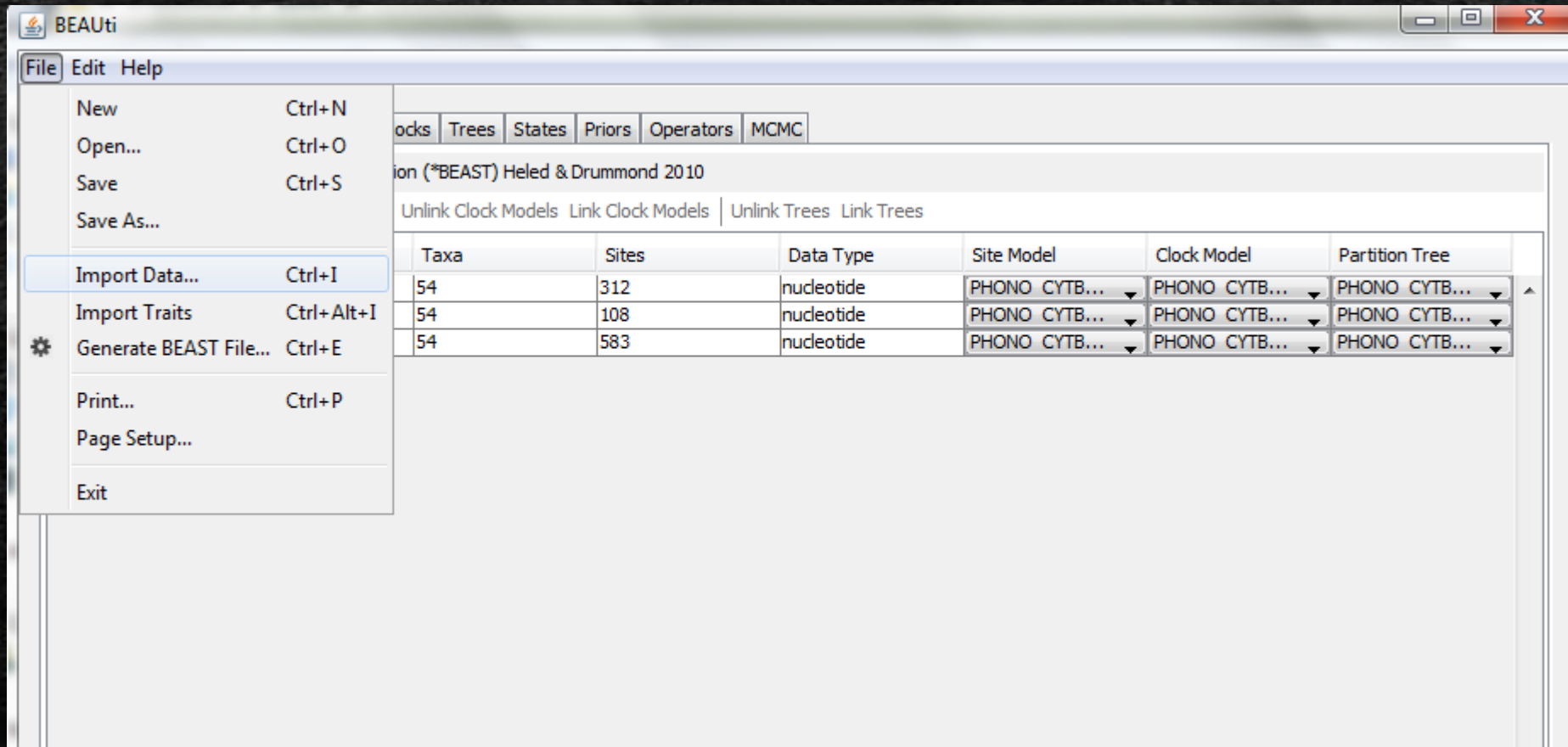
Modeller	Bezier 1mL	Harmonik 1mL	LBF (Bezier)	Sıralama (Bezier)	Model Olasılığı
Model 1	-4862.85	-4791.29	-4.54	2	0.087
Model 2	-4860.58	-4795.88	0.00	1 (en iyi)	0.844
Model 3	-4863.08	-4794.53	-5.00	3	0.069
Model 4	-4887.25	-4803.22	-53.34	4	0.000

Phonochorion populasyonlarında ortak ataya varma zamanının hesaplanması

- **Beast:** http://beast.bio.ed.ac.uk/Main_Page
- **Tracer:** <http://tree.bio.ed.ac.uk/software/tracer/>
- **Veri Seti:** *Phonochorion* Clade III veri seti
- **Amaç:** Doğu Karadenizde yayılan *Phonochorion* cinsine ait III nolu soy hattının farklı demografik büyüme modelleri altında ortak ataya varma zamanlarının (Tmrca) hesaplanması
 - 1. Model: Standard birleşme modeli altında Tmrca'nın hesaplanması
 - 2. Model: Demografik büyüme birleşme modeli altında Tmrca'nın hesaplanması
 - 3. Model: Bayesian Skyline Plot yöntemi ile zaman içinde populasyon büyüklüğü değişiminin modellenmesi ve Tmrca'nın hesaplanması

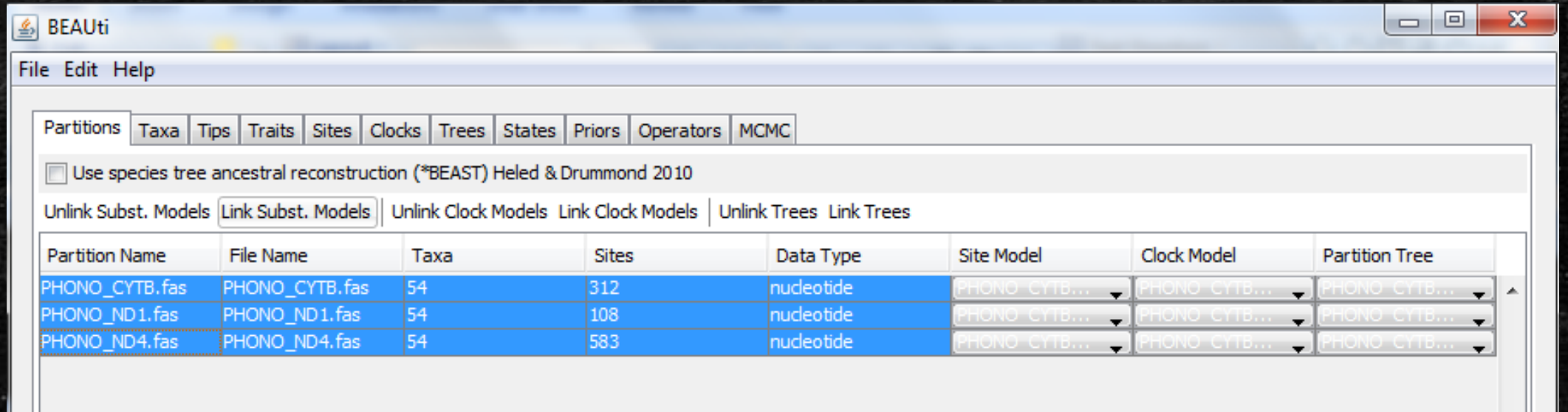
Beast Analizi için Beauti ile xml dosyası oluşturma: Veri Girişi

Beauti programını açın ve **Import Data** seçeneği ile gen dizilerine ait dosyaları yükleyin.

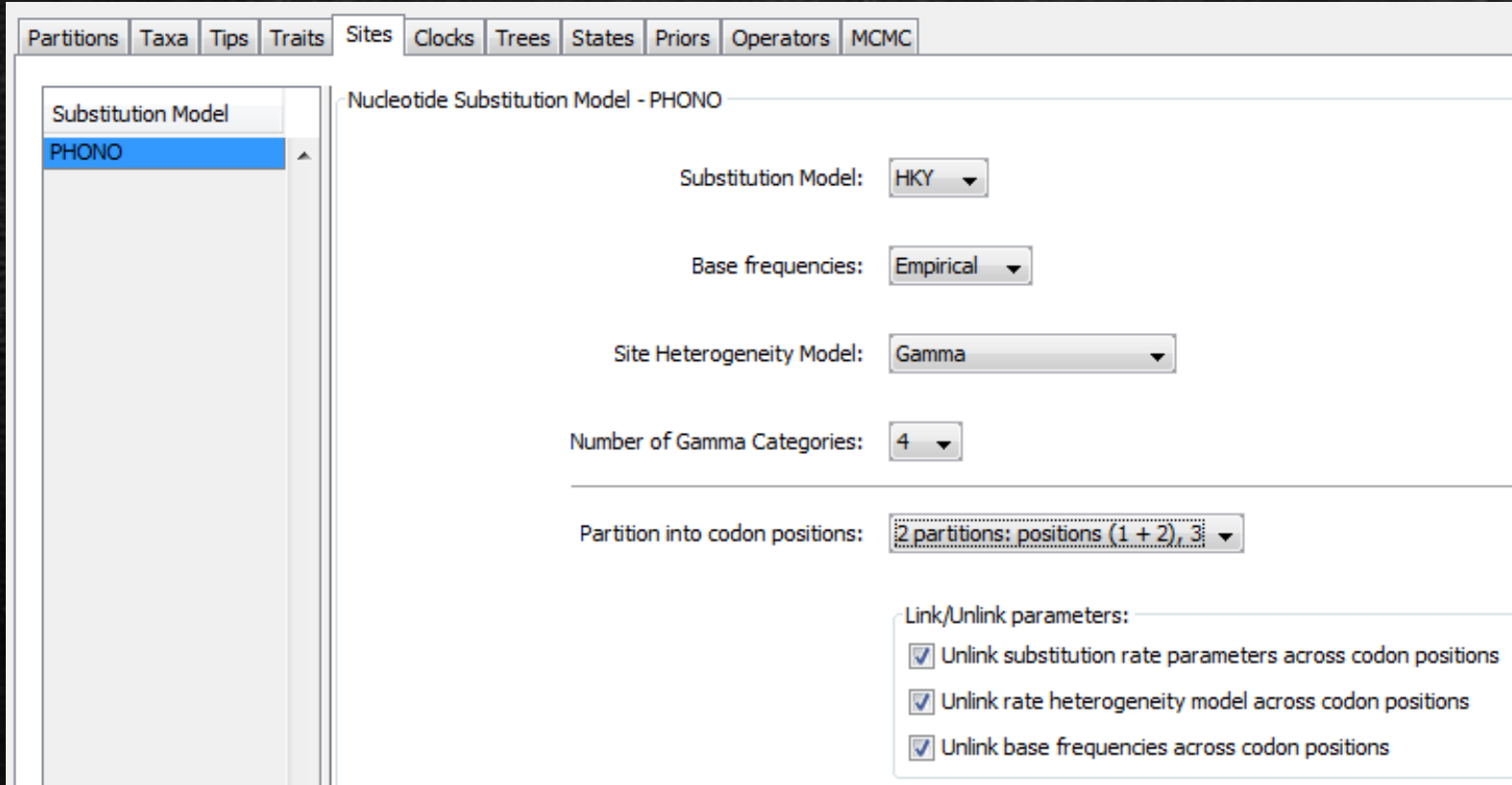


Beast Analizi için Beauti ile xml dosyası oluşturma: Farklı genleri bağlama çözme

Her üç genide seçin ve **Link** seçeneğini kullanarak **Subst. Models**, **Clock Models** ve **Trees** fonksiyonları için her üç genide bir birine bağlayın. Böylece her üç gende tek bir mutasyon, moleküler saat ve ağaç modeli ile tanımlanabilecek. Eğer **unlinked** seçeneğini kullanırsanız her bir gen için ayrı ayrı modeller kullanacaksınız anlamına gelir.



Beast Analizi için Beati ile xml dosyası oluşturma: Mutasyon modeli



The screenshot shows the Beati software interface with the 'Sites' tab selected. The 'Nucleotide Substitution Model - PHONO' settings are displayed. The 'Substitution Model' is set to 'HKY', 'Base frequencies' to 'Empirical', and 'Site Heterogeneity Model' to 'Gamma'. The 'Number of Gamma Categories' is set to '4'. The 'Partition into codon positions' is set to '2 partitions: positions (1 + 2), 3'. The 'Link/Unlink parameters' section has three checkboxes: 'Unlink substitution rate parameters across codon positions', 'Unlink rate heterogeneity model across codon positions', and 'Unlink base frequencies across codon positions', all of which are checked.

Partitions Taxa Tips Traits Sites Clocks Trees States Priors Operators MCMC

Substitution Model

PHONO

Nucleotide Substitution Model - PHONO

Substitution Model: HKY

Base frequencies: Empirical

Site Heterogeneity Model: Gamma

Number of Gamma Categories: 4

Partition into codon positions: 2 partitions: positions (1 + 2), 3

Link/Unlink parameters:

- ☒ Unlink substitution rate parameters across codon positions
- ☒ Unlink rate heterogeneity model across codon positions
- ☒ Unlink base frequencies across codon positions

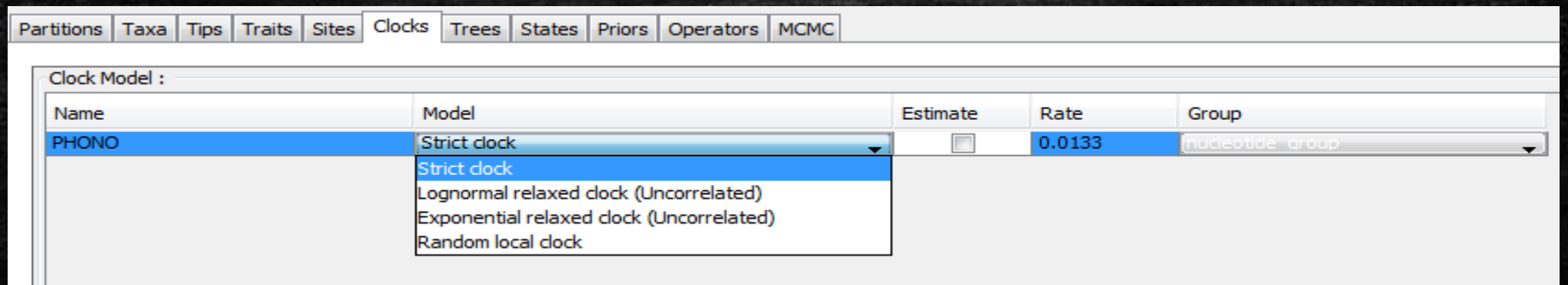
Sites sayfasına gelerek gen bölgesine ilişkin uygun mutasyon modelini giriniz.

Söz konusu bölge mtDNA'ya ilişkin protein kodlayan genler olduğu için modeli kodonlara göre böleriz. Böylece 3. kodonun 1 ve 2. kodona göre daha yüksek görece mutasyon hızı göz önüne alınmış olur.

Beast Analizi için Beauti ile xml dosyası oluşturma: Moleküler saat modeli

Clocks sayfasına gelerek gen bölgesine ilişkin uygun moleküler saat modelini giriniz.

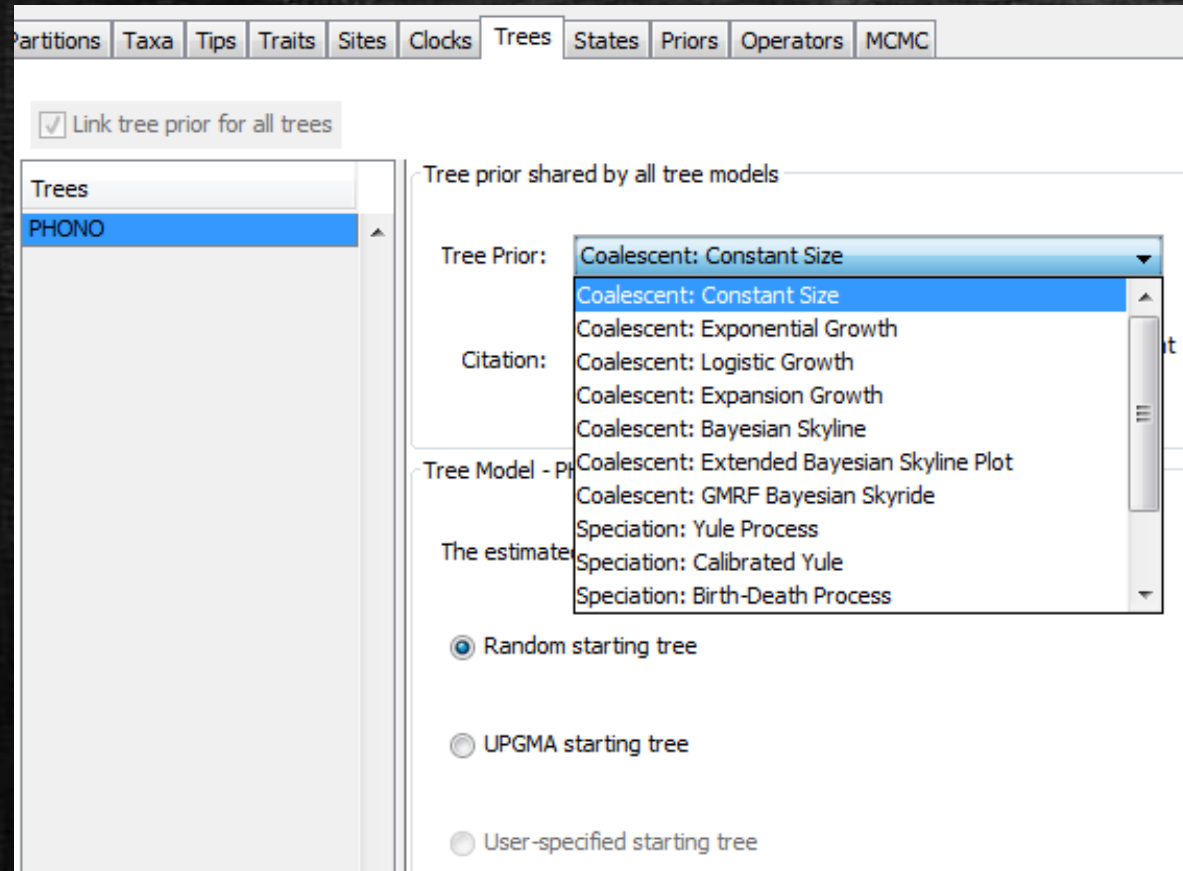
Bir çok seçenek olmasına rağmen şimdilik sabit moleküler saati kullanacağız. Saat hızını milyon yılda bölge ve soy başına “0.0133” olarak belirleriz. Bu mtDNA için oldukça tutarlı olan milyon yılda sekanslar arası % 2.66’lık farklılaşma oranına denk gelir.



Name	Model	Estimate	Rate	Group
PHONO	Strict clock	☐	0.0133	nucleotide group

- Strict clock
- Lognormal relaxed clock (Uncorrelated)
- Exponential relaxed clock (Uncorrelated)
- Random local clock

Beast Analizi için Beauti ile xml dosyası oluşturma: Ağaç modeli



Trees sayfasına gelerek veriye uygun ağaç modelini giriniz.

Birleşme ve türleşme süreçlerine bağlı farklı ağaç modelleri olduğunu göreceksiniz. Eğer **populasyon birleşme tabanlı** yöntemleri, **türler** üzerinden analiz yapıyorsanız **türleşme modellerini** kullanmalısınız.

Şimdilik Standart Birleşme modeli olan **Coalescent: Constant Size** modelini seçiniz.

Beast Analizi için Beati ile xml dosyası oluşturma: Ön olasılık dağılımları

The screenshot shows the BEAST software interface with the 'Priors' tab selected. The 'Priors for model parameters and statistics' table lists various parameters and their default priors. A dialog box titled 'Prior for Parameter CP1+2.mu' is open, showing the 'Prior Distribution' set to 'Uniform' with an 'Initial value' of 1.0, 'Upper' bound of 1000, and 'Lower' bound of 0.0.

Parameter	Prior	B...	Description
CP1+2.kappa	* LogNormal [1, 1.25], initial=2	[...]	HKY transition-transversion parameter for codon positions 1 & 2
CP3.kappa	* LogNormal [1, 1.25], initial=2	[...]	HKY transition-transversion parameter for codon position 3
CP1+2.mu	! Uniform infinite bounds, initial=1	[...]	relative rate parameter for codon positions 1 & 2
CP3.mu	! Uniform infinite bounds, initial=1	[...]	relative rate parameter for codon position 3
CP1+2.alpha	* Exponential [0.5], initial=0.5	[...]	gamma shape parameter for codon positions 1 & 2
CP3.alpha	* Exponential [0.5], initial=0.5	[...]	gamma shape parameter for codon position 3
treeModel.rootHeight	* Using Tree Prior in [0, ∞]	[...]	root height of the tree
constant.popSize	* 1/x, initial=0.2	[...]	coalescent population size parameter

Link parameters into a phylogenetic hierarchical model
* Marked parameters currently have a default prior distribution. You should check the 'Priors' tab in the BEAST GUI.

Data: 54 taxa, 3 partitions; Fix clock rate to 0.0133 in nucleotide_group;

Priors sayfasından ardıl olasılık dağılımlarını elde etmede kullanacağımız ön olasılık dağılımlarını modifiye edebiliriz.

Bayesian analizlerinin can damarı burasıdır. Doğru ön olasılık dağılımları etkin ardıl olasılık dağılımlarına yol açar.

Bugün BEAST'in önerdiği ön olasılık dağılımlarını kullanacağız. Sadece BEAST kodon bölgelerinin görece hızları için uygun bir ön olasılık dağılımı öneremediği için bu parametreleri tanımlayan **CP1+2.mu** ve **CP3.mu** parametrelerinin ön olasılık dağılımını alt ve üst sınırları 0 – 1000 olan tekdüze (uniform) bir dağılım olarak ayarlayın.

Beast Analizi için Beati ile xml dosyası oluşturma: MCMC zincirinin ayarlanması

ocks Trees States Priors Operators MCMC

Length of chain: 1000000

Echo state to screen every: 100

Log parameters every: 100

File name stem: CLADE_III_CONS

☒ Add .txt suffix

Log file name: CLADE_III_CONS.log.txt

Trees file name: CLADE_III_CONS.trees.txt

☒ Create operator analysis file:

Analysis file name: CLADE_III_CONS.ops.txt

Generate BEAST File...

MCMC sayfasından zincirimizin uzunluğunu ve hangi aralıkta bu zincirden örnek seçeceğimizi belirleriz.

Bugünkü analizlerimiz için zincir uzunluğu 1 mily. olacak (length of chain) ve her 100 zincirde bir örnekleme (Log paramters every) alacağız. Böylece ardıl olasılık dağılımlarını belirlemek için toplamda 10 bin değer kullanmış olacağız. Not: 1 mily. luk bir zincir son derece kısadır ve normal bir analiz en az 10 mily. zincirden oluşmalıdır. Bir çok durumda ise 50 veya 100 mily.'luk zincirler gereklidir.

Zincir uzunluğunu ayarladıktan analize bir isim veririz (File name) ve sonra Generate Beast File tuşuna basarak "xml" dosyası oluşturmuş oluruz.

Beast Analizi ve Sonuç Dosyaları

Beast v1.7.exe dosyasını açın ve **choose file** seçeneğini kullanarak hazırlamış olduğunuz **xml** dosyasını girip analizi başlatmak için “**run**”a basın.

MCMC zinciri tamamlandığında BEAST size üç adet dosya verecektir:

- İsim.log.txt:** Bütün tahminlediğimiz parametre değerlerin ardıl olasılık dağılımlarını içeren dosya
- İsim.ops.txt:** Zincirin çalışması ile ilgili operatör dosyası
- İsim.trees.txt:** Bütün tahminlediğimiz gen ağaçlarını içeren dosya

Beast Analizi ve Sonuç Dosyaları

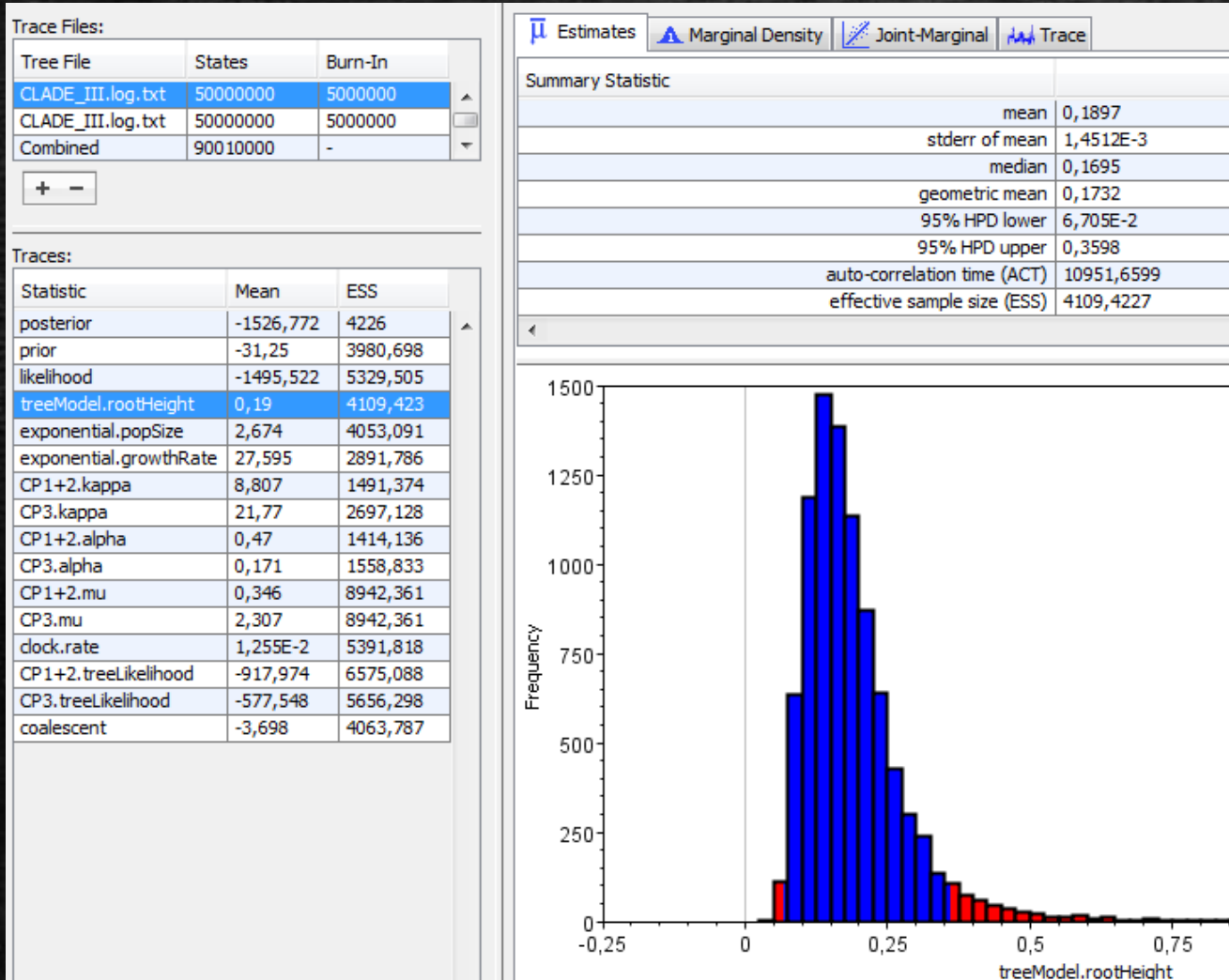
Bugün yapacaklarımızda bizi sadece “log” dosyası yani parametre tahminlerinin ardıl olasılıklarını içeren dosya ilgilendiriyor.

Ardıl olasılık dağılımlarını görebilmek için **tracer** programını açınız ve “import trace file” seçeneğini kullanarak **isim.log.txt** dosyasını açınız.

Tracer, MCMC analizinin tahminlediği bütün ardıl olasılık dağılımları ile ilgili “özet istatistikleri” ve en önemlisi bütün parametrelere ilişkin “tarama iz”lerini verecektir.

Analizimizin ne kadar sağlıklı yürüdüğünü ve parametre tahminlerimizin ne kadar iyi olduğunu buradaki değerleri inceleyerek anlayabiliriz.

Tipik bir Tracer dosyası



Yanda üssel büyüme birleşme modeli altında Tmrca'nın (treeModel.rootHeight) hesaplanmasına ilişkin analizin sonuçları özetlenmektedir.

Görüleceği gibi tracer programı her bir tahminlenen parametrenin ardıl olasılık dağılımları ile ilgili “Ort.” “Std.hata” ve “%95 Güven Aralıkları” gibi bir çok önemli istatistiği hesaplar.

Dikkat edilmesi gereken en önemli üç husus güven aralıklarının dar, dağılımın tek tepeli ve iz dosyalarının “beyaz sesi” andırıyor olması.

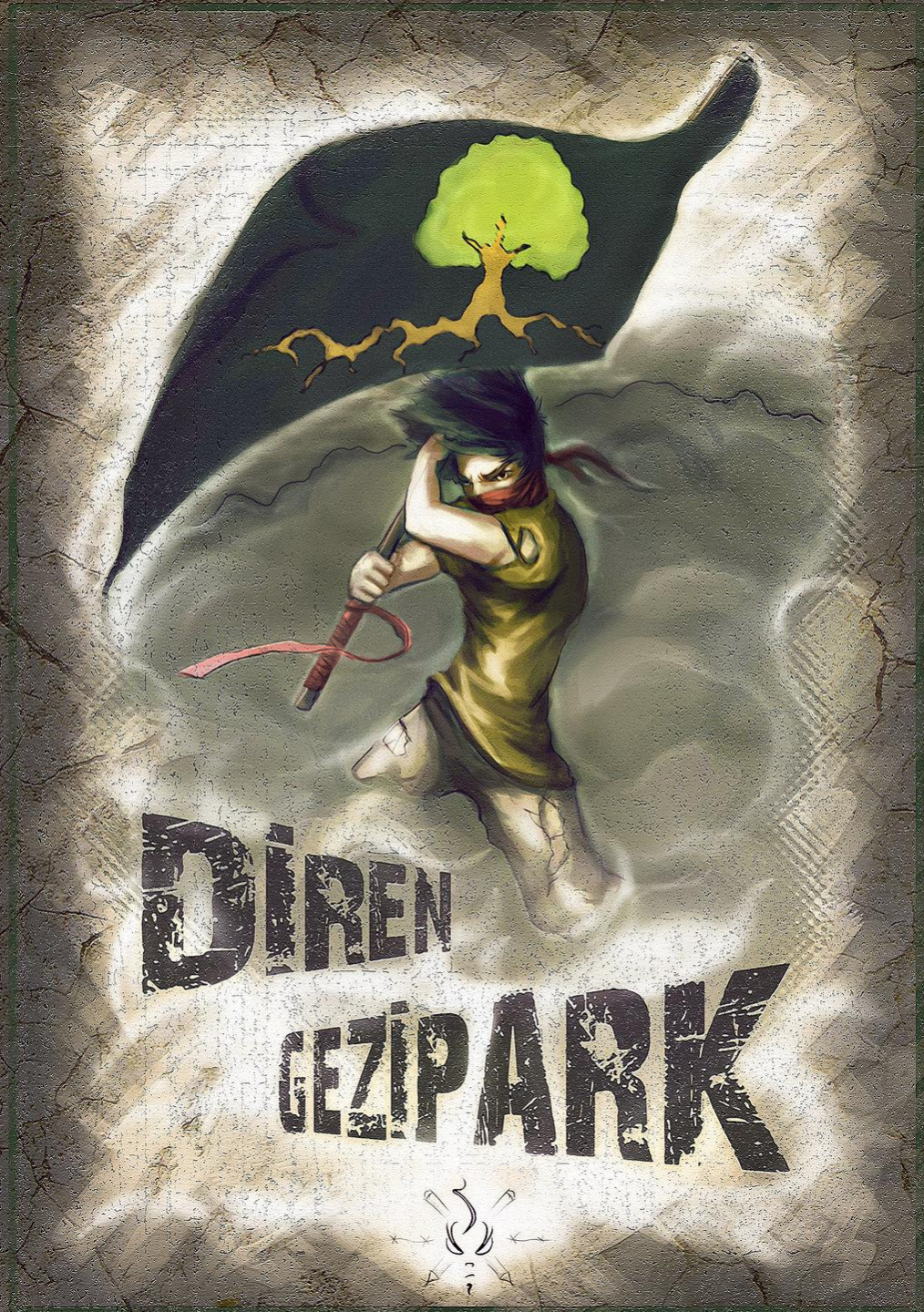
Phonochorion populasyonlarında ortak ataya varma zamanının hesaplanması

- *Phonochorion* Clade III veri setinin Beauti programı ile açarak her bir model için farklı xml dosyası hazırlayın.
- Bütün analizlerde HKY+G mutasyon modeli kullanılacaktır. Veriyi kodon bölgelerine göre [(1+2)+3] bölmeyi unutmayınız!!!
- 1. Model: Standard birleşme modeli altında Tmrca'nın hesaplaması
 - Trees sayfasında ağaç modeli olarak “Coalescent: Constant Size”ı seçiniz
- 2. Model: Demografik büyüme birleşme modeli altında Tmrca'nın hesaplanması
 - Trees sayfasında ağaç modeli olarak “Coalescent: Exponential Growth”u seçiniz
- 3. Model: Bayesian Skyline Plot
 - Trees sayfasında ağaç modeli olarak “Extended Bayesian Skyline Plot”u seçiniz

Phonochorion populasyonlarında ortak ataya varma zamanının hesaplanması

- Beast analizi sonucunda elde ettiğiniz “log” dosyalarını tracer altında inceleyiniz.
- Model 1 için *Tmrca* (treeModel.rootHeight) parametresine ilişkin ortalama, medyan, %95 güven aralıklarını ve etkin örneklem sayısını (ESS) aşağıdaki çizelgeye not ediniz.
- Model 2 için *Tmrca* değerlerinin yanında üssel büyüme (exponential growth rate) parametresine ilişkin değerleri çizelgeye not ediniz.
- Model 3 için *Tmrca* değerlerinin yanında çıkan “csv” dosyasını kullanarak excel’de etkin populasyon büyüklüğünün (*N_{ef}*)’in zaman içindeki değişimine ilişkin medyan ve %95 güven aralıklarını gösteren çizgi grafiği oluşturun.

	Tmrca			Üssel Büyüme
	Model 1	Model 2	Model 3	Model 2
ortalama				
ort. standard hata				
medyan				
95% HPD alt				
95% HPD üst				
etkin örneklem büyüklüğü (ESS)				



D İ R E N
E V R İ M

D İ R E N
O D T Ü

D İ R E N
T Ü R K İ Y E