Phylogenetic Inference via Categorical Spectra

Categorical Time-series of Protein Amino-acid Sequences

Sumaira Qureshi[1] Roger Brown[2]

Phylomania: Hobart Tasmania. October 2009

Sumaira Qureshi

Insert portrait here.

Stimulus

Ph.D. Thesis topic for Sumaira (Part-II)

- Submitted for examination: July-2009
- Examiners reports received a few days ago.
- Outcome not known at this time.

What is a Categorical Time-series ?

- Consider the sequence: ABFDEHBFDEHBCDKAJGDE
- Assign categories:
  $\alpha = \{A, H\}, \beta = \{B, J\}, \gamma = \{C, F, G\}, \delta = \{D\}, \varepsilon = \{E, K\}$
- List by categories: $\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon$
- Fourier techniques are useful for establishing categories for sequences with inherent periodicities.

Categorical Time-series

What is a Categorical Time-series ?

- Consider the sequence: ABFDEHBFDEHBCDKAJGDE
- Assign categories:
  $\alpha = \{A, H\}, \beta = \{B, J\}, \gamma = \{C, F, G\}, \delta = \{D\}, \varepsilon = \{E, K\}$
- List by categories: $\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon$
- Fourier techniques are useful for establishing categories for sequences with inherent periodicities.

What is a Categorical Time-series ?

- Consider the sequence: ABFDEHBFDEHBCDKAJGDE
- Assign categories:
  $\alpha = \{A, H\}, \beta = \{B, J\}, \gamma = \{C, F, G\}, \delta = \{D\}, \varepsilon = \{E, K\}$
- List by categories: $\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon$
- Fourier techniques are useful for establishing categories for sequences with inherent periodicities.

Categorical Time-series

What is a Categorical Time-series ?

- Consider the sequence: ABFDEHBFDEHBCDKAJGDE
- Assign categories:
  $\alpha = \{A, H\}, \beta = \{B, J\}, \gamma = \{C, F, G\}, \delta = \{D\}, \varepsilon = \{E, K\}$
- List by categories: $\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon$
- Fourier techniques are useful for establishing categories for sequences with inherent periodicities.

What is a Categorical Time-series ?

- Consider the sequence: ABFDEHBFDEHBCDKAJGDE
- Assign categories:
  $\alpha = \{A, H\}, \beta = \{B, J\}, \gamma = \{C, F, G\}, \delta = \{D\}, \varepsilon = \{E, K\}$
- List by categories: $\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon\,\alpha\beta\gamma\delta\varepsilon$
- Fourier techniques are useful for establishing categories for sequences with inherent periodicities.

Previous Applications of Fourier Analysis to Protein Sequences

Far too many to mention them all, but two notable examples are:

- The Atchley group: Prof. William R. Atchley, North Carolina State University
- The Resonance Recognition Method (RRM), under continuing development by Irena Cosic *et al*.

Almost all previous studies have two features in common:

1. Each amino-acid is assigned a numerical value derived from a table physio-chemical properties, such as electropositivity, hydrophobicity etc.
2. A frequency analysis method that focuses on the local *instantaneous* frequency. (*e.g.* Burg and Lewis correlation matrices)

Previous Applications of Fourier Analysis to Protein Sequences

Far too many to mention them all, but two notable examples are:

- The Atchley group: Prof. William R. Atchley, North Carolina State University
- The Resonance Recognition Method (RRM), under continuing development by Irena Cosic *et al*.

Almost all previous studies have two features in common:

1. Each amino-acid is assigned a numerical value derived from a table physio-chemical properties, such as electropositivity, hydrophobicity etc.
2. A frequency analysis method that focuses on the local *instantaneous* frequency. (*e.g.* Burg and Lewis correlation matrices)

Previous Applications of Fourier Analysis to Protein Sequences

Far too many to mention them all, but two notable examples are:

- The Atchley group: Prof. William R. Atchley, North Carolina State University
- The Resonance Recognition Method (RRM), under continuing development by Irena Cosic *et al*.

Almost all previous studies have two features in common:

1. Each amino-acid is assigned a numerical value derived from a table physio-chemical properties, such as electropositivity, hydrophobicity etc.
2. A frequency analysis method that focuses on the local *instantaneous* frequency. (*e.g.* Burg and Lewis correlation matrices)

Features of this approach

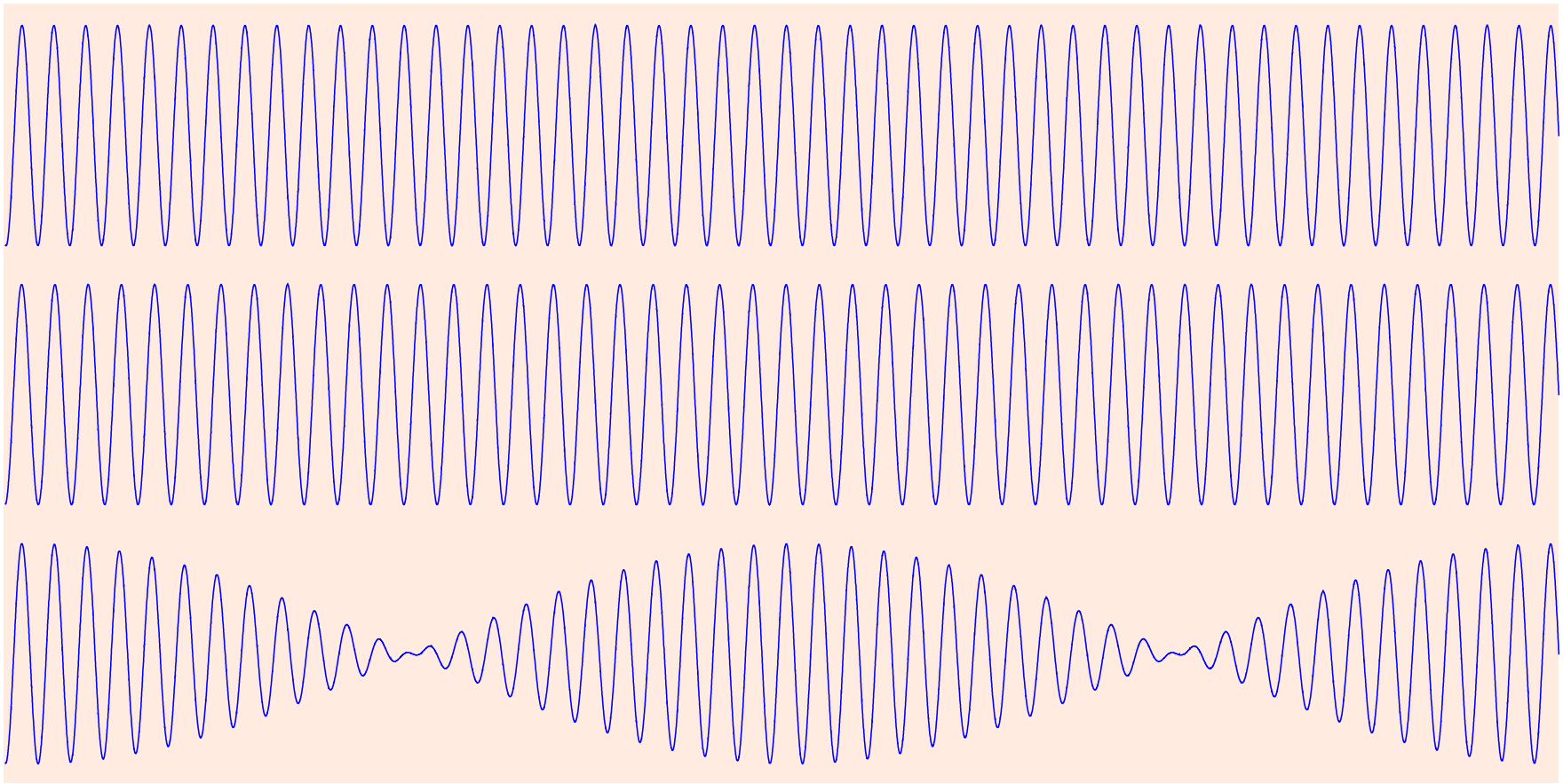
1. Numerical values are assigned to each amino-acid type in a manner to optimize some chosen metric of the sequence. No consideration is given to any of the physical or chemical properties of any particular residue type. Furthermore , the assigned numerical values are not constant but depend on spatial frequency. In other words a different spectrum of values is assigned to each amino-acid for each protein sequence.
2. Frequency analysis techniques adopt a global rather than local perspective. i.e. Long-range spatial frequency rather than the instantaneous frequency.

What is *"instantaneous"* frequency ?

$$\sin(\alpha+\delta)+\sin(\alpha-\delta) = 2\sin(\alpha)\cos(\delta)$$

Illustration of acoustic *"beats"*

$$\sin(\alpha+\delta)+\sin(\alpha-\delta) = 2\sin(\alpha)\cos(\delta)$$



The Canadian, Delhousie Group

- Examining Protein Structure and Similarities by Spectral Analysis Technique (2006)
  Krista Collins, Hong Gu, Chris Field
- Spectral analysis for categorical time series: Scaling and the spectral envelope (1993)
  Stoffer, D S. and Tyler, D E. and McDougall, A J. (Biometrika)

Variations on the Stoffer method

- Scalar algorithm

$$E(\omega) = \min_{\phi, |\beta\rangle} \frac{1}{N_a} \sum_{\{t\}} \left(\sin(\omega t + \phi) - \beta_{\tau(t)}\right)^2$$

- Complex algorithm

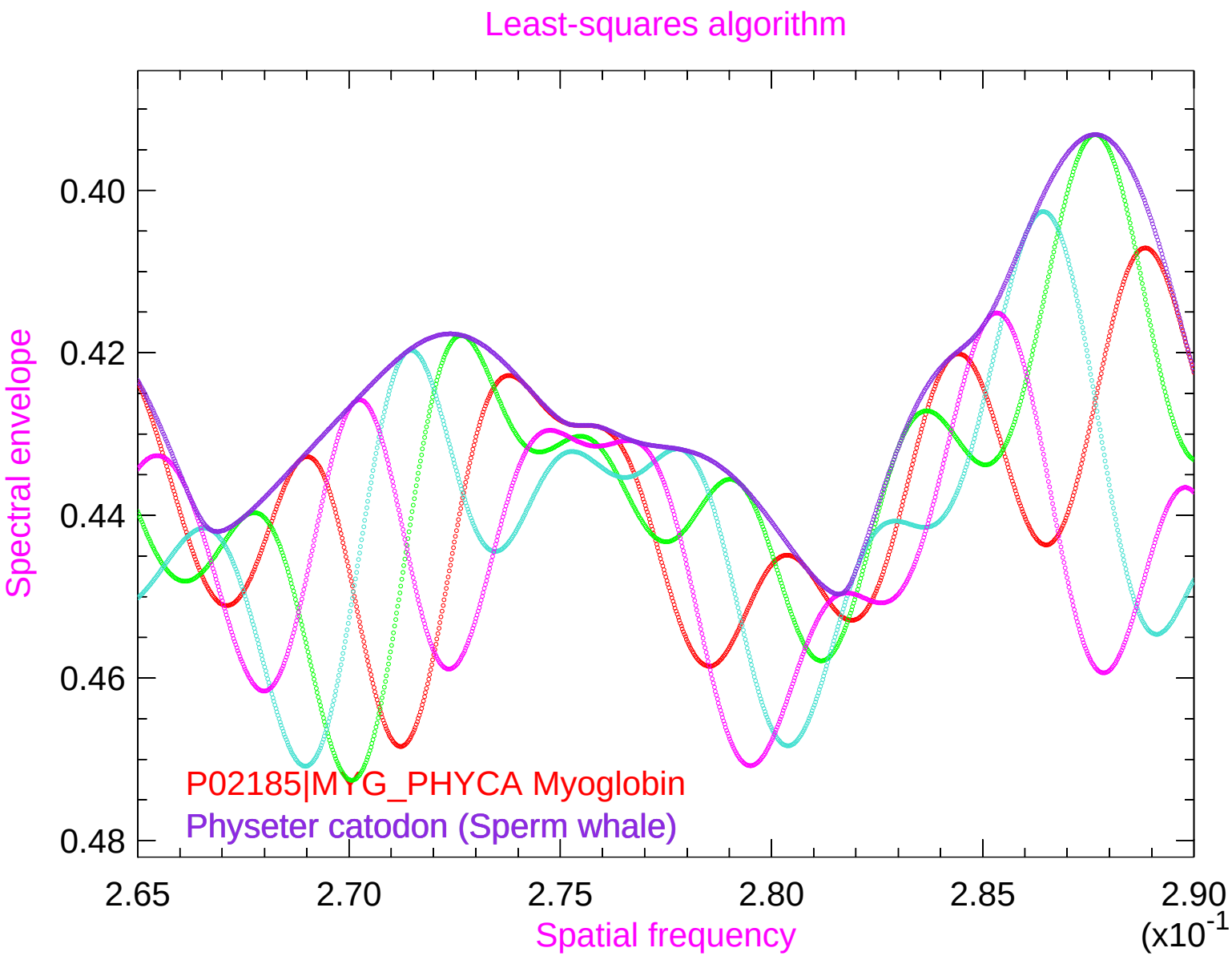
$$E(\omega) = \min_{|\beta\rangle} \frac{1}{N_a} \sum_{\{t\}} \left(e^{i\omega t} - \beta^*_{\tau(t)}\right)\left(e^{-i\omega t} - \beta_{\tau(t)}\right)$$

Sperm-whale Myoglobin Protein Sequence

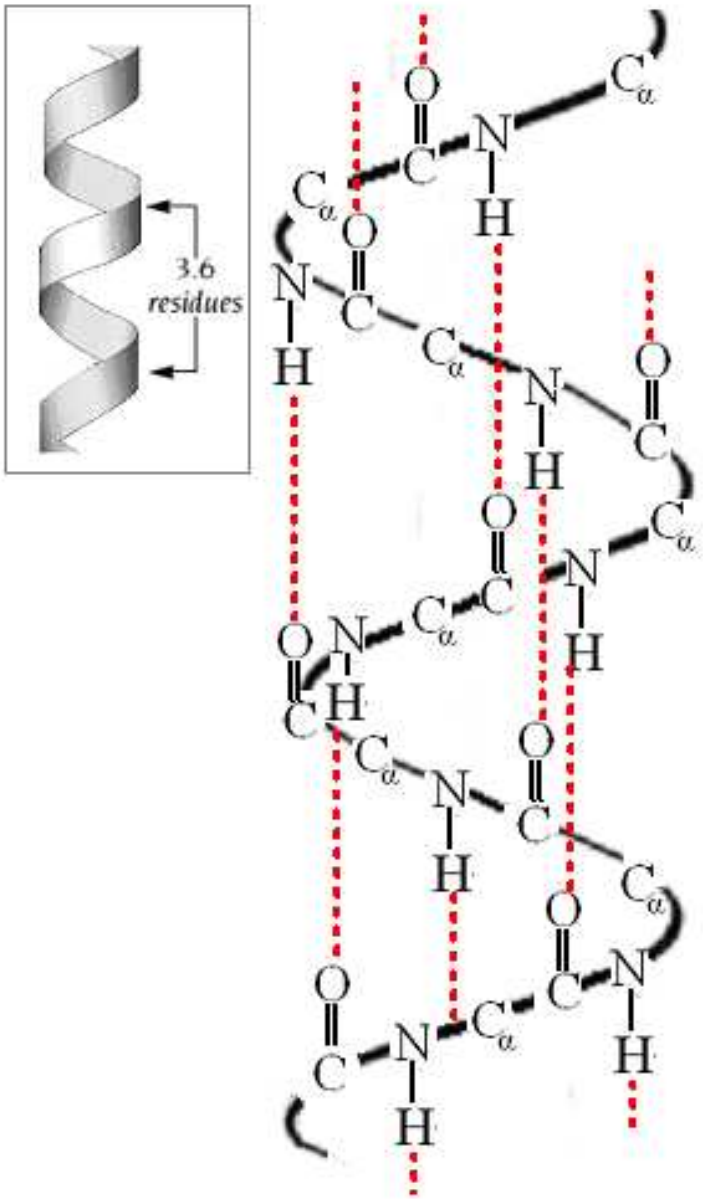
>P02185|MYG_PHYCA Myoglobin - Physeter catodon (Sperm whale)
(Physeter macrocephalus).

VLSEGEWQLVLHVWAKVEADVAGHGQDILIRLFKSHPETLEKFDRFKHLKTE
AEMKASEDLKKHGVTVLTALGAILKKKGHHEAELKPLAQSHATKHKIPIKYL
EFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG

Scalar Spectral Envelope



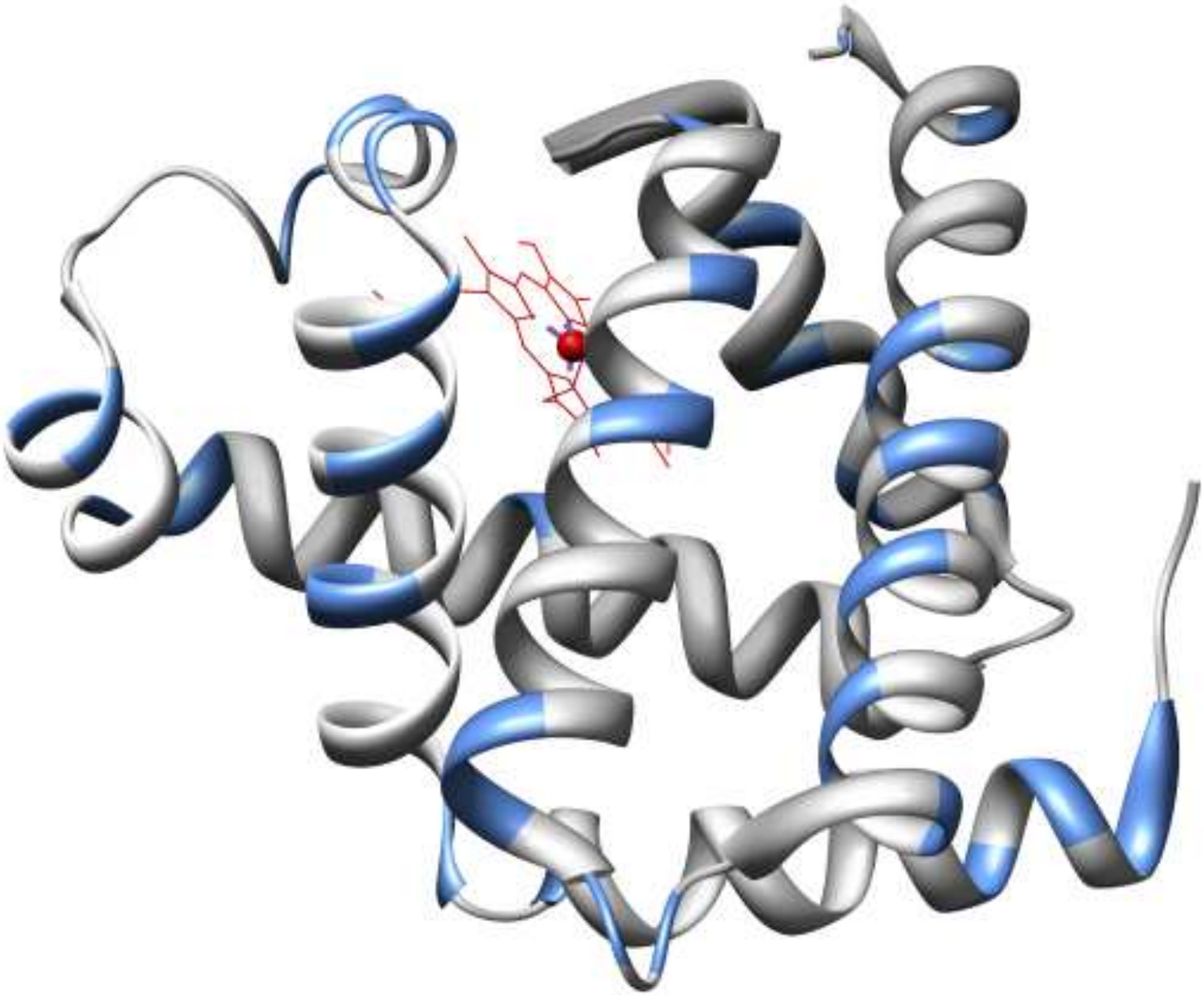
α–helix Structure



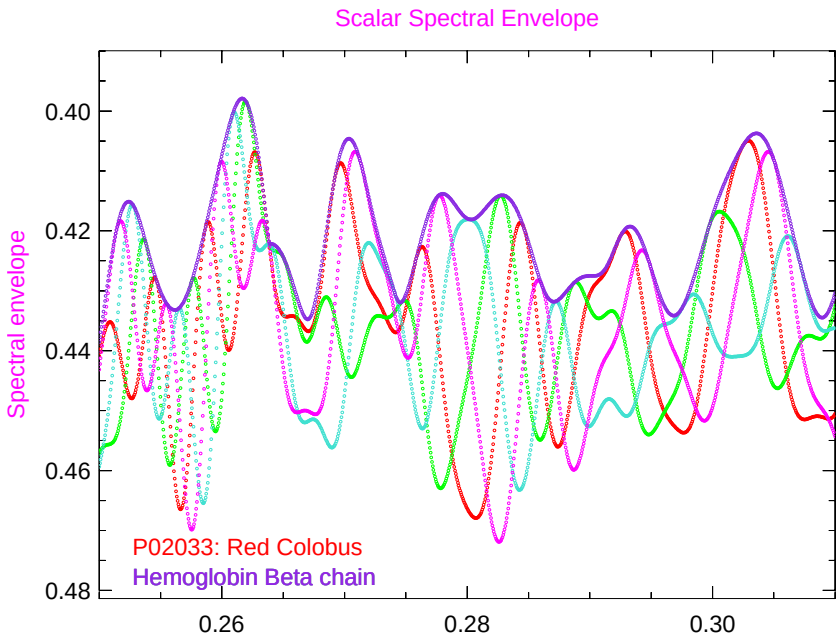
Spatial wavelength - Spatial frequency

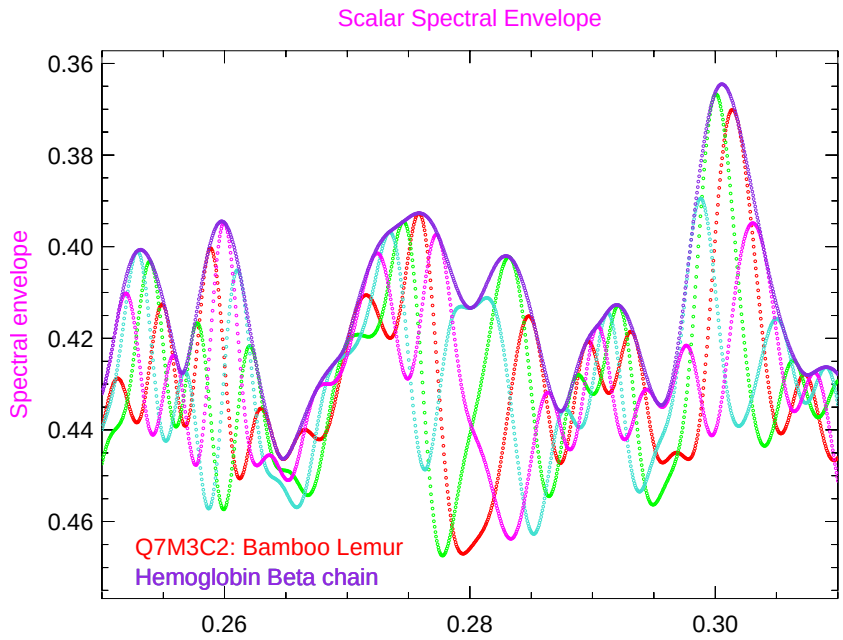
$$1/3.6 \quad = \quad 0.277777\ldots$$

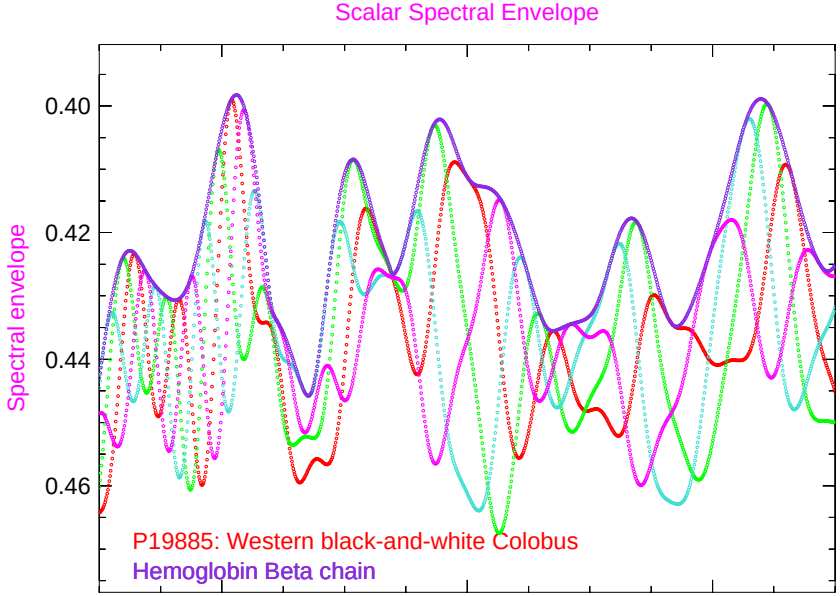
Myoglobin Structure

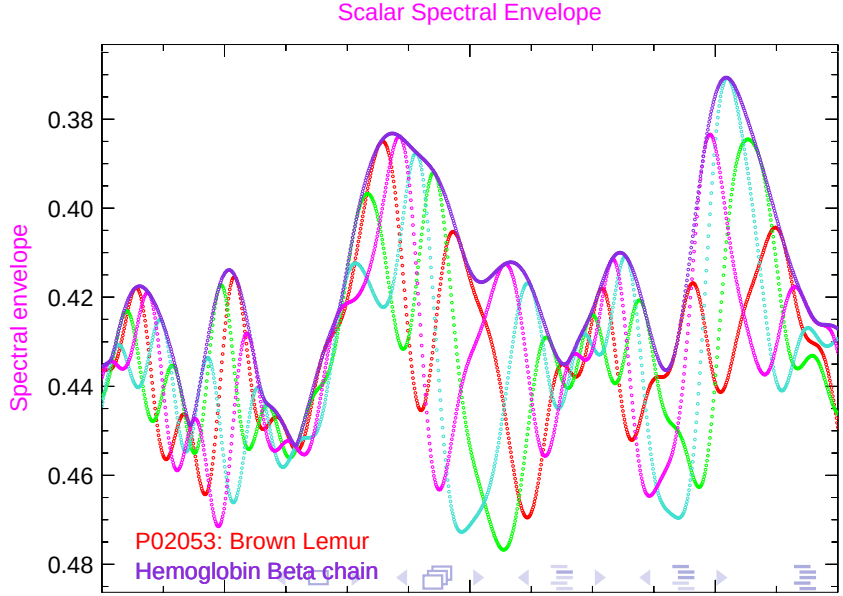


Phylogenetic inference from Categorical Spectra



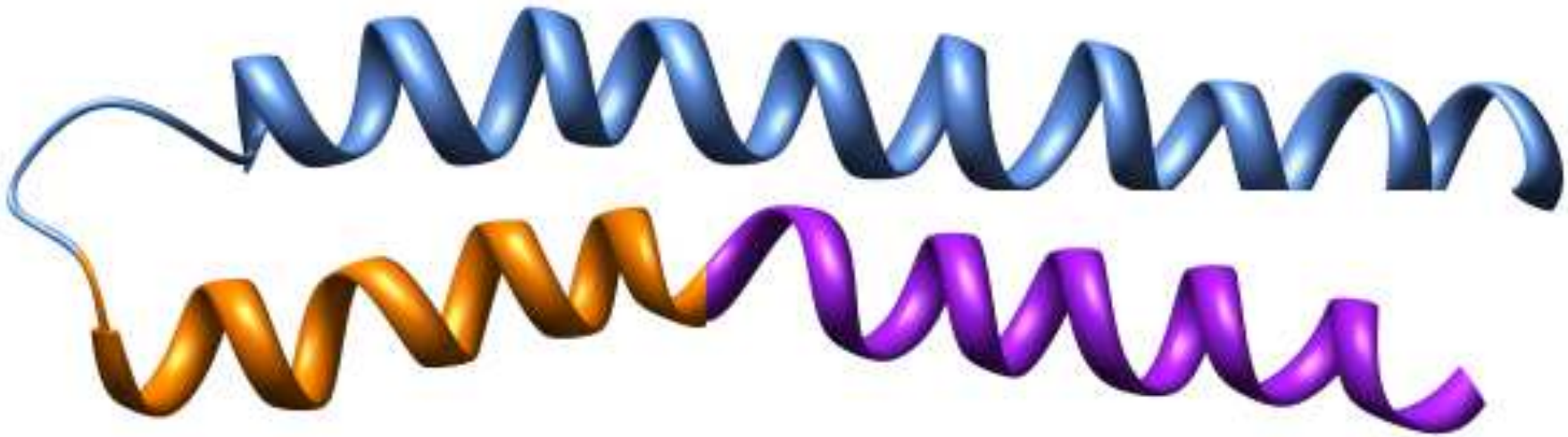






Rotor part of the ATP-synthase molecule

An ATP-synthase helix pair

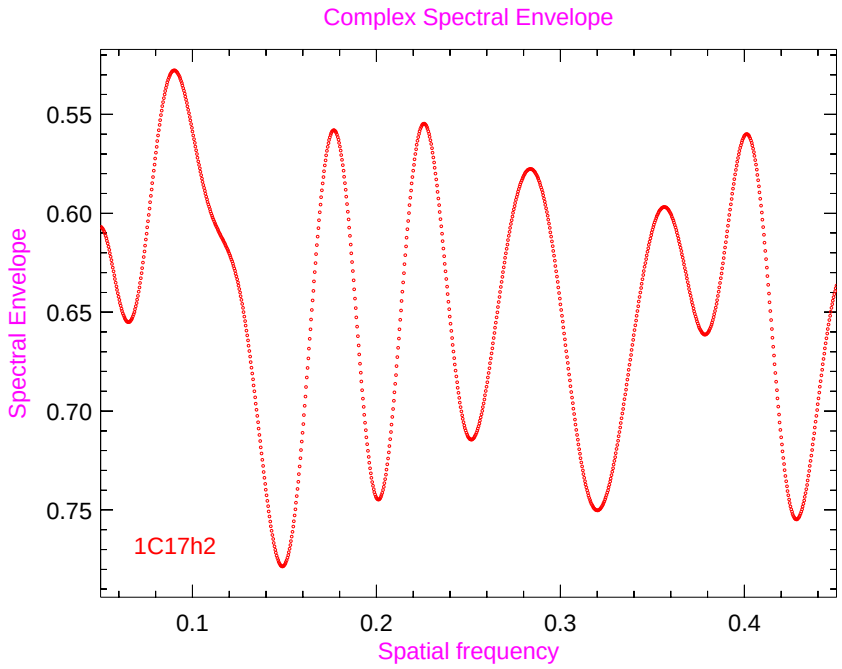


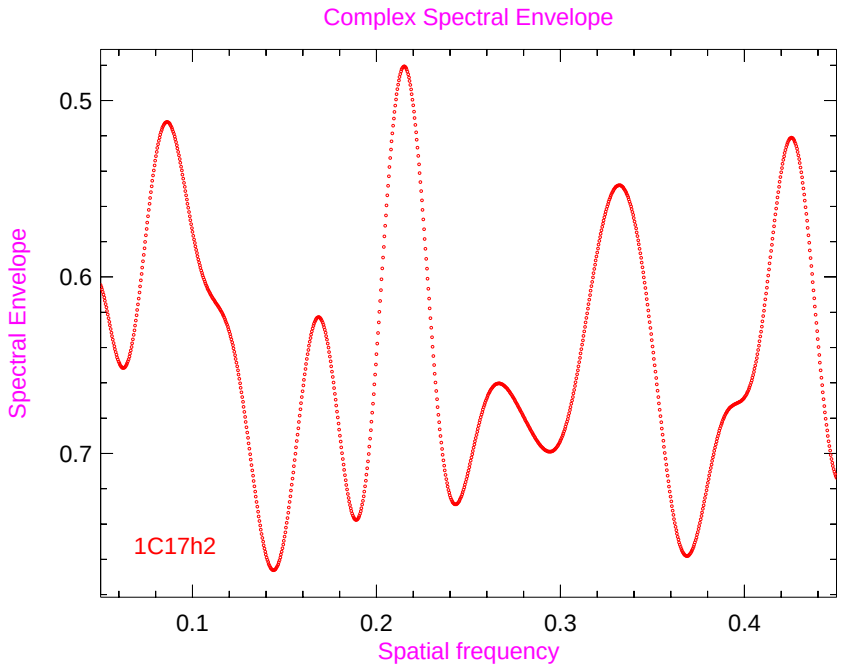
ATP-Synthase gapped sequences

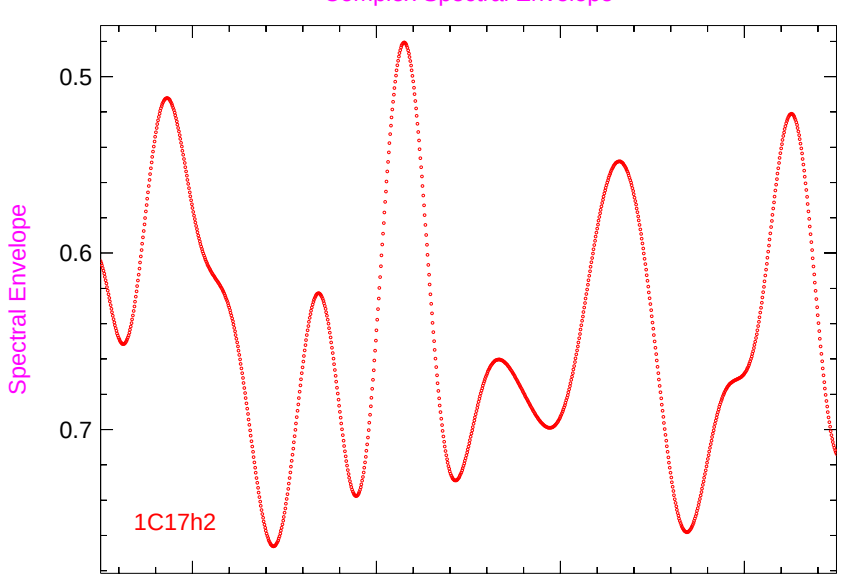
LIPLLR TQFFIVMGLV		DAIPMIAVGL GLYVMFAVA
LIPLLR TQFFIVMGLV	-	DAIPMIAVGL GLYVMFAVA
LIPLLR TQFFIVMGLV	–	DAIPMIAVGL GLYVMFAVA
LIPLLR TQFFIVMGLV	—	DAIPMIAVGL GLYVMFAVA

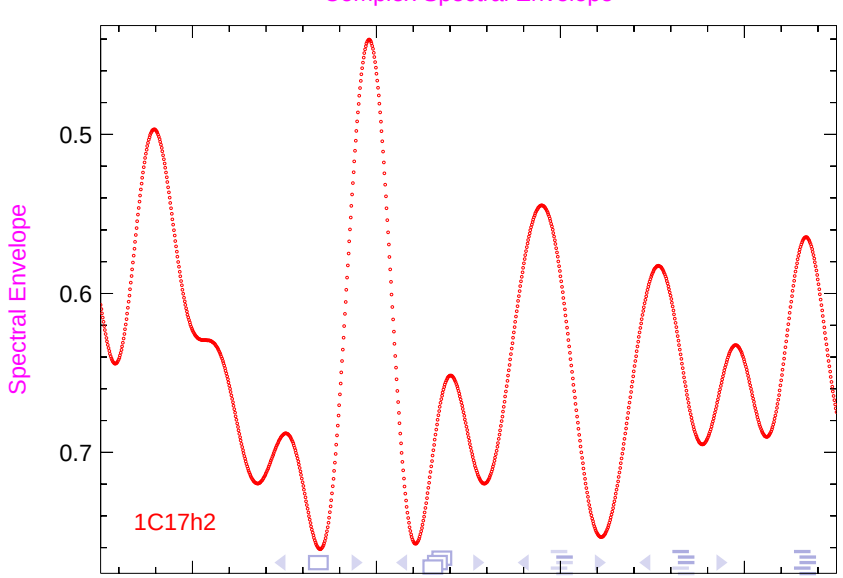
Table: Amino-acid sequences of ATP-synthase outer helices.

Complex Categorical spectra

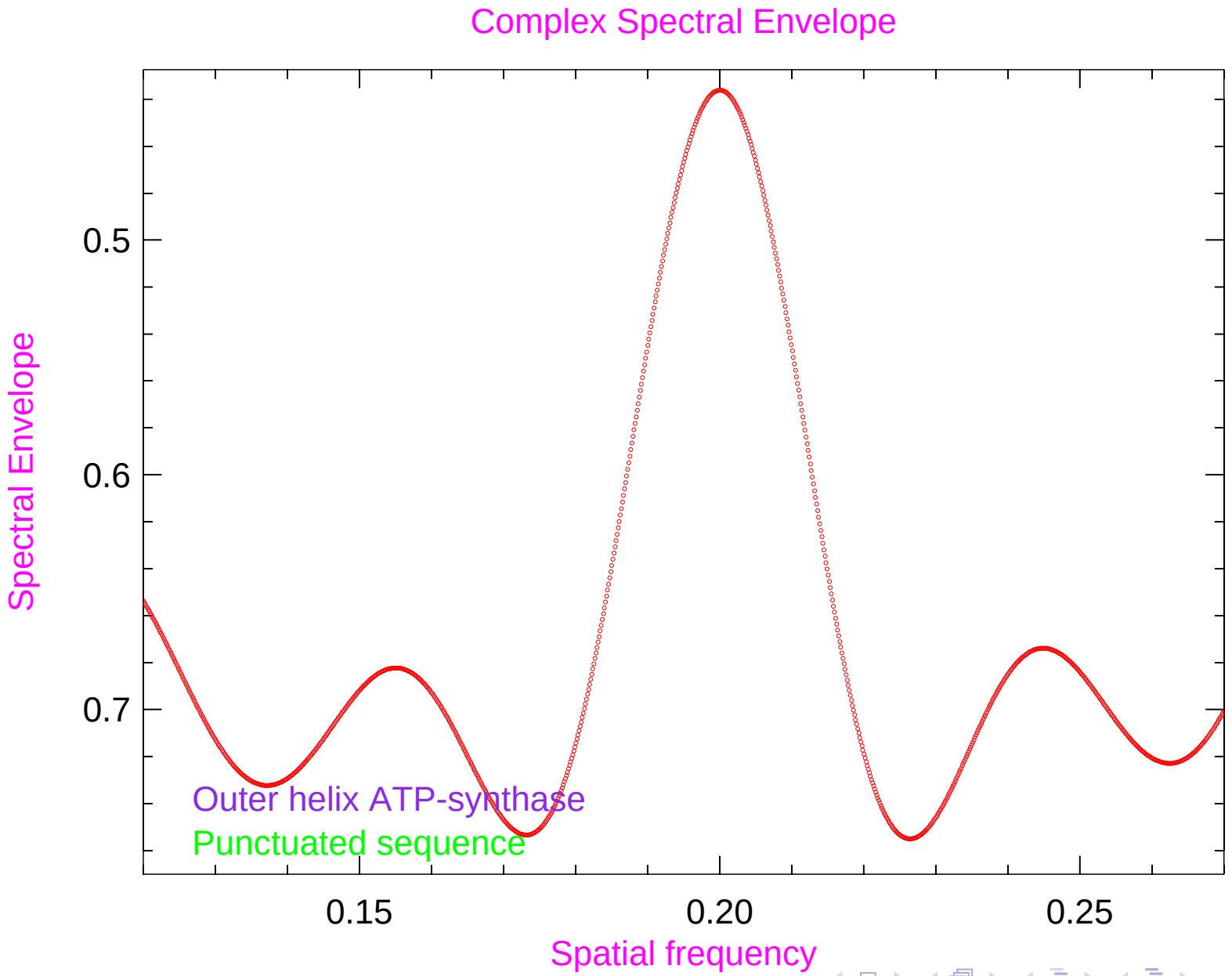








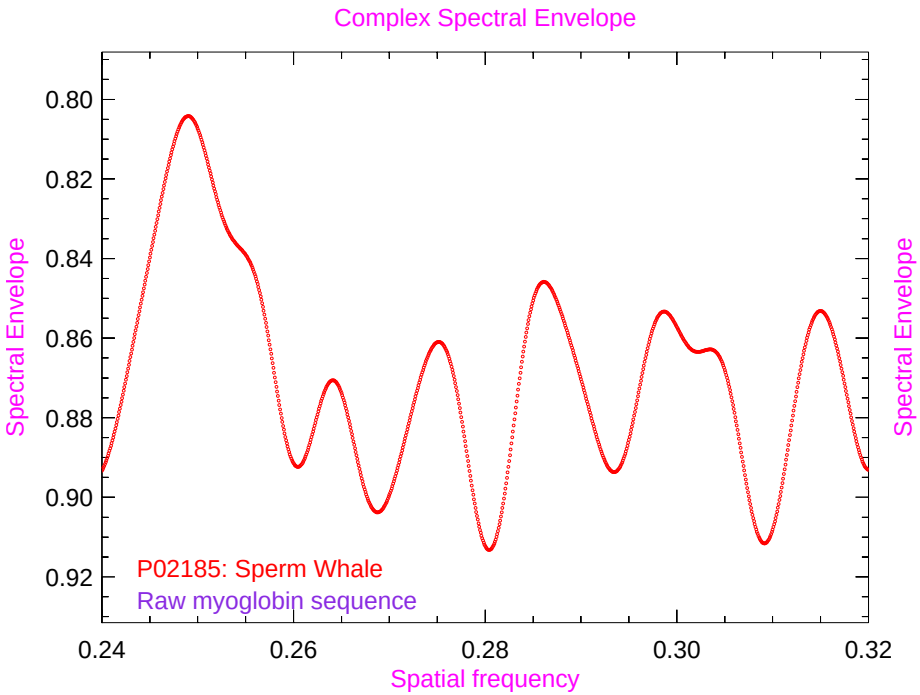
Fractional spacing

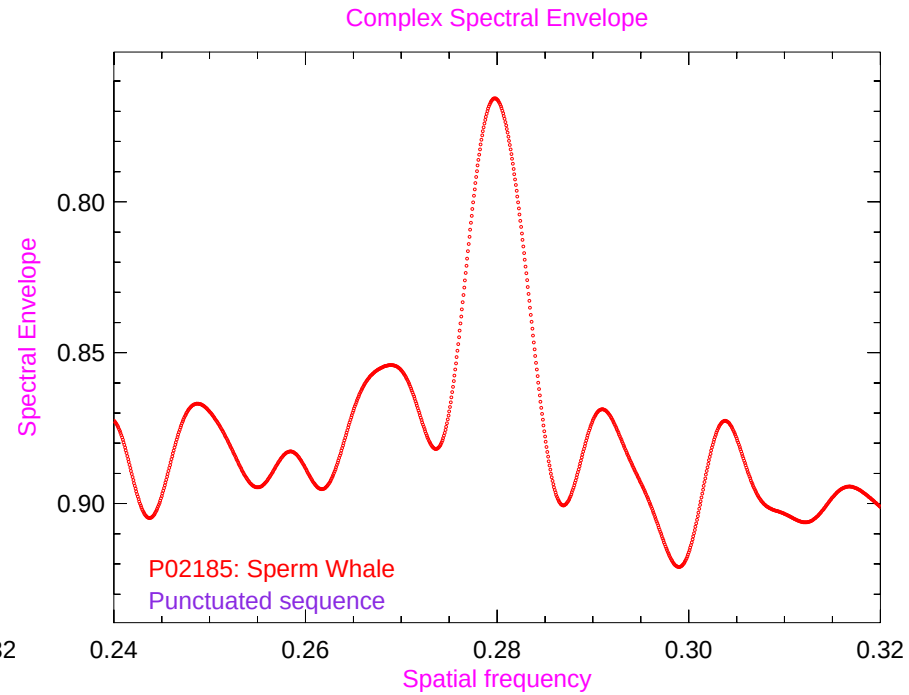


Punctuated Sperm-whale myoglobin sequence

$>$ *P*02185|*MYG* − *HYCA Myoglobin*
VL
$<$ 2.51235 $>$SEGEWQLVLHVWAKVEA
$<$ 3.48089 $>$DVAGHGQDILIRLFKS
$<$ 0.96800 $>$HPETLEKFDRFKHLK
$<$ 3.18381 $>$TEAEMKA
$<$ 1.75174 $>$SEDLKKHGVTVLTALGAILKKKGHHEAE
$<$ 1.32164 $>$LKPLAQSHATKHKI
$<$ 2.41373 $>$PIKYLEFISEAIIHVLHSRHPGDF
$<$ 0.94676 $>$GADAQGAMNKALELFRKDIAAKYKELGYQG

Raw and punctuated sperm-whale complex spectra





Distance measure between two categorical spectra

If $\mathscr{S}$ is a residue sequence of length N, consider $X(\mathscr{S};t) = \beta_{\tau(t)} = \langle Y(t) | \beta(\omega) \rangle$. $X(\mathscr{S};t)$ is a complex (or real) sequence associated with the categorical sequence $\mathscr{S}$

$$\Lambda(\mathscr{S}_1, \mathscr{S}_2) = \frac{1}{N_a} \int_{\omega=0}^{\omega=\infty} W(\omega) \sum_{\{t\}} \| X(\mathscr{S}_1;t) - X(\mathscr{S}_2;t) \|^2 \, d\omega$$

arrives at the following equation for ε:

$$C_1 D_{22}^{T} \sin\psi_{22}^{T} = k_2 \sin\varepsilon + C_1 D_{22}^{A} \sin(\psi_{22}^{A} + \delta) - C_2 \sin\varepsilon + C_3 E_{22}^{air} \sin(\Upsilon_{22}^{air} + \delta) \quad (2)$$

where $C_1 = 3\rho_w(1+k_2')/(5\rho_e\bar{H}) = 0.9590\,\mathrm{m}^{-1}$, $C_2 = 3\rho_w h_2(1+k_2')/(5\rho_e) = 4.752 \times 10^{-2}$, and $C_3 = 3(1+k_2')/(5g\rho_e\bar{H}) = 9.445 \times 10^{-5}\,\mathrm{Pa}^{-1}$; k_2 and k_2' are the potential Love and loading numbers (constants that describe how Φ^S depends on the Earth tide and on Earth loading[12]), h_2 the displacement Love number (describing the vertical displacement of the Earth tide[12]), ρ_w and ρ_e are the densities of sea water and Earth, g the gravitational acceleration, and $\bar{H}$ for M_2 equals 8.137 cm. D_{22}^{A}, ψ_{22}^{A} are the coefficients of the altimeter-derived ocean tide (which were derived assuming that $\varepsilon = 0$ and therefore required the small C_2 correction term), and E_{22}^{air}, Υ_{22}^{air} are the coefficients of the lunar barometric tide[21,22]. Angle δ represents an additional phase offset in the ocean plus load tide and the air plus load tide, induced by anelastic response to Earth loading. Because equation (2) is fairly insensitive to δ, we hereafter set it to zero. (If δ were as large as 1°, its neglect would affect the estimated ε by only 0.07°. Further, there is no physical reason to expect that $\delta \gg \varepsilon$; theoretical calculations based on various Earth models[23,24] suggest that $\delta < 0.02°$.) It is straightforward to show that ε is also insensitive to the other adopted constants required in C_1, C_2 and C_3.

Equation (2) is analogous to the simpler equation (8) of Zschau[1]. Zschau was able to write an additional equation for the in-phase component ($k_2 \cos\varepsilon$) and thereby estimated both k_2 and ε (at least in principle; in practice the ocean models available to him were too inaccurate for this). Here, because of our use of altimetry, we cannot estimate k_2 independently of h_2. Alternatively, we may use the in-phase components to place bounds on δ. This must be examined in more detail elsewhere, but the bounds turn out to be weak and, given the lack of any systematic in-phase difference in Fig. 1, the result is not significantly different from zero.

For the tidal coefficients D_{22}^{A}, ψ_{22}^{A}, D_{22}^{T}, ψ_{22}^{T}, we adopt the means of the appropriate models listed in Table 1. For the tracking and altimeter standard errors, we use 0.024 and 0.036 cm, respectively; the latter is simply the listed value for the one model and ignores averaging as all three altimeter models use essentially the same data. Solution of equation (2) then yields $\varepsilon = 0.16° \pm 0.09°$. This value of ε implies[13] that friction in the Earth's body tide consumes 83 ± 45 gigawatts of tidal power. For comparison, this is ~3% of the ocean's dissipation, and, according to Platzman's estimate[22], about eight times the atmosphere's dissipation. If Q^{-1} is the body tide's specific dissipation function, then the effective tidal $Q = 1/\tan\varepsilon$ is found to be 370, with approximate confidence limits of (200,800).

This estimate of Q helps fill an important observational gap between seismic modes of period less than 1 hour and the Chandler wobble of period 14 months; even measurements with relatively wide confidence bounds may help clarify previous disparate theoretical and/or indirect estimates. For example, one indirect approach[2] infers tidal Q from free-oscillation data, arguing that the body tide and the ${}_0S_2$ free-oscillation mode should, aside from a slight frequency dependence, display similar Q (although Bostrom[3] suggests that the Earth's rotation may reduce the tidal Q). Munk[4], in fact, adopted a free-oscillation Q of 350 (now thought to be too low[5]) to estimate a solid-Earth tidal dissipation essentially equivalent to our observational estimate. Other theoretical estimates of complex k_2 Love numbers[6] suggest bounds on tidal Q of (90, 500), with a preference around 210, whereas another[7] k_2 value suggests a Q of 126, which is outside our (1σ) observational range. Finally, Zschau[1], adopting a single absorption-band model between seismic and Chandler wobble periods, arrives at 'most probable' estimates for ε of 0.21° and solid-Earth dissipation of 120 GW. Within their error bounds, our observations are consistent with Zschau's values and with the assumed smooth frequency dependence between seismic and Chandler periods. □

Received 19 January; accepted 8 May 1996.

1. Zschau, J. in *Space Geodesy and Geodynamics* (eds Anderson, A. & Cazenave, A.) 315–344 (Academic, London, 1986).
2. Lagus, P. L. & Anderson, D. L. *Phys. Earth planet. Inter.* **1,** 505–510 (1968).
3. Bostrom, R. C. *EOS* **76,** F60 (1995).
4. Munk, W. *Q. Jl R. astr. Soc.* **9,** 352–375 (1968).
5. Tanimoto, T. *Geophys. Res. Lett.* **17,** 669–672 (1990).
6. Wahr, J. & Bergen, Z. *Geophys. J. R. astr. Soc.* **87,** 633–668 (1986).
7. Dehant, V. *Phys. Earth planet. Inter.* **49,** 97–116 (1987).
8. Lambeck, K. *Phil. Trans. R. Soc. Lond.* **A287,** 545–594 (1977).
9. Ray, R. D. in *The Oceans* (eds. Majumdar, S. K. *et al.*) 171–185 (Pennsylvania Academy of Science, Easton, 1994).
10. Fu, L.-L. *et al. J. geophys. Res.* **99,** 24369–24381 (1994).
11. Cohen, S. C. & Smith, D. E. *J. geophys. Res.* **90,** 9217–9220 (1985).
12. Munk, W. H. & MacDonald, G. J. F. *The Rotation of the Earth* (Cambridge Univ. Press, 1960).
13. Platzman, G. W. *Rev. Geophys. Space Phys.* **22,** 73–84 (1984).
14. Cartwright, D. E. & Ray, R. D. *J. geophys. Res.* **96,** 16897–16912 (1991).
15. Schrama, E. J. O. & Ray, R. D. *J. geophys. Res.* **99,** 24799–24808 (1994).
16. Marshall, J. A. *et al. J. geophys. Res.* **100,** 25331–25352 (1995).
17. Le Provost, C., Bennett, A. F. & Cartwright, D. E. *Science* **267,** 639–642 (1995).
18. Andersen, O. B., Woodworth, P. L. & Flather, R. A. *J. geophys. Res.* **100,** 25261–25282 (1995).
19. Lerch, F. *Bull. Géod.* **65,** 44–52 (1991).
20. Wunsch, C. & Stammer, D. *J. geophys. Res.* **100,** 24895–24910 (1995).
21. Haurwitz, B. & Cowley, A. D. *Pure appl. Geophys.* **102,** 193–222 (1973).
22. Platzman, G. W. *Pure Appl. Geophys.* **137,** 1–33 (1991).
23. Zschau, J. in *Tidal Friction and the Earth's Rotation* (eds Brosche, P. & Sündermann, J.) 62–94 (Springer, New York, 1978).
24. Pagiatakis, S. D. *Geophys. J. Int.* **103,** 541–560 (1990).
25. Ray, R. D., Sanchez, B. & Cartwright, D. E. *EOS* **75,** 108 (1994).
26. Eanes, R. J. *EOS* **75,** 108 (1994).
27. Lerch, F. J. *et al.* NASA Tech. Memo. 104555 (Goddard Space Flight Center, Greenbelt, 1992).
28. Tapley, B. D., Schutz, B. E., Eanes, R. J., Ries, J. C. & Watkins, M. M. in *Cotnributions of Space Geodesy to Geodynamics: Earth Dynamics* (eds Smith, D. E. & Turcotte, D.) 147–174 (Am. geophys. Union, Washington, 1993).
29. Gill, A. *Atmosphere-Ocean Dynamics* (Academic, New York, 1982).
30. Le Provost, C., Genco, M. L., Lyard, F., Vincent, P. & Canceil, P. *J. geophys. Res.* **99,** 24777–24797 (1994).
31. Schwiderski, E. W. *Mar. Geol.* **6,** 219–265 (1983).
32. Parke, M.E. & Hendershott, M. C. *Mar. Geol.* **3,** 379–407 (1980).

ACKNOWLEDGEMENTS. We thank C. F. Yoder and especially G. W. Platzman for useful discussions, and C. Wunsch for his frequency-wavenumber spectra.

CORRESPONDENCE should be addressed to R.D.R. (e-mail: ray@nemo.gsfc.nasa.gov).

The guinea-pig is not a rodent

Anna Maria D'Erchia*†, Carmela Gissi*†, Graziano Pesole‡, Cecilia Saccone*§ & Ulfur Arnason†

* Dipartimento di Biochimica e Biologia Molecolare, Università di Bari, 70125 Bari, Italy
† Department of Evolutionary Molecular Systematics, University of Lund, Sölvegatan 29, S-22362 Lund, Sweden
‡ Dipartimento di Biologia DBAF, Università della Basilicata, 30100 Potenza, Italy
§ Centro di Studio sui Mitocondri e Metabolismo Energetico, Consiglio Nazionale delle Ricerche, 70125 Bari, Italy

In 1991 Graur *et al.* raised the question of whether the guinea-pig, *Cavia porcellus*, is a rodent[1]. They suggested that the guinea-pig and myomorph rodents diverged before the separation between myomorph rodents and a lineage leading to primates and artiodactyls. Several findings have since been reported, both for and against this phylogeny, thereby highlighting the issue of the validity of molecular analysis in mammalian phylogeny. Here we present findings based on the sequence of the complete mitochondrial genome of the guinea-pig, which strongly contradict rodent monophyly. The conclusions are based on the cumulative evidence provided by orthologically inherited genes and the use of three different analytical methods, none of which joins the guinea-pig with myomorph rodents. In addition to the phylogenetic conclusions, we also draw attention to several factors that are important for the validity of phylogenetic analysis based on molecular data.

The order Rodentia is by far the most speciose mammalian order, comprising 1,814 species and 29 families. It is traditionally divided into three suborders: Sciuromorpha (squirrels), Myomorpha (mice, rats), and Hystricomorpha (porcupines, guinea-pigs).

Journal of Mammalian Evolution, Vol. 4, No. 2, 1997

Are Guinea Pigs Rodents? The Importance of Adequate Models in Molecular Phylogenetics

Jack Sullivan[1,2] and David L. Swofford[1]

The monophyly of Rodentia has repeatedly been challenged based on several studies of molecular sequence data. Most recently, D'Erchia *et al.* (1996) analyzed complete mtDNA sequences of 16 mammals and concluded that rodents are not monophyletic. We have reanalyzed these data using maximum-likelihood methods. We use two methods to test for significance of differences among alternative topologies and show that (1) models that incorporate variation in evolutionary rates across sites fit the data dramatically better than models used in the original analyses, (2) the mtDNA data fail to refute rodent monophyly, and (3) the original interpretation of strong support for nonmonophyly results from systematic error associated with an oversimplified model of sequence evolution. These analyses illustrate the importance of incorporating recent theoretical advances into molecular phylogenetic analyses, especially when results of these analyses conflict with classical hypotheses of relationships.

KEY WORDS: *inconsistency; maximum likelihood; molecular systematics; rodents; rate heterogeneity.*

INTRODUCTION

The assertions made in several molecular phylogenetic studies (Graur *et al.*, 1991; Li *et al.*, 1992; Ma *et al.*, 1993) have led to the growing acceptance of the conclusion that the order Rodentia is not monophyletic, in spite of the facts that these data sets essentially provide no significant refutation of the classical hypothesis (e.g., Hasegawa *et al.*, 1992; Cao *et al.*, 1994), and other molecular studies actually support rodent monophyly (Martignetti and Brosius, 1993; Porter *et al.*, 1996). Recently, D'Erchia *et al.* (1996) suggested that their phylogenetic analyses of complete mtDNA sequences of 16 mammalian species firmly establish that the guinea pig is not a rodent, based on its placement as a sister taxon to a clade containing Lagomorpha, Carnivora, Primates, Perissodactyla, and Artiodactyla (including cetaceans), rather than in a clade with mouse and rat. They claim that this placement both is consistent across phylogenetic reconstruction methodologies and is supported by "very significant" bootstrap values. Because nonmonophyly of the rodents would imply a remarkable amount of convergence in mor-

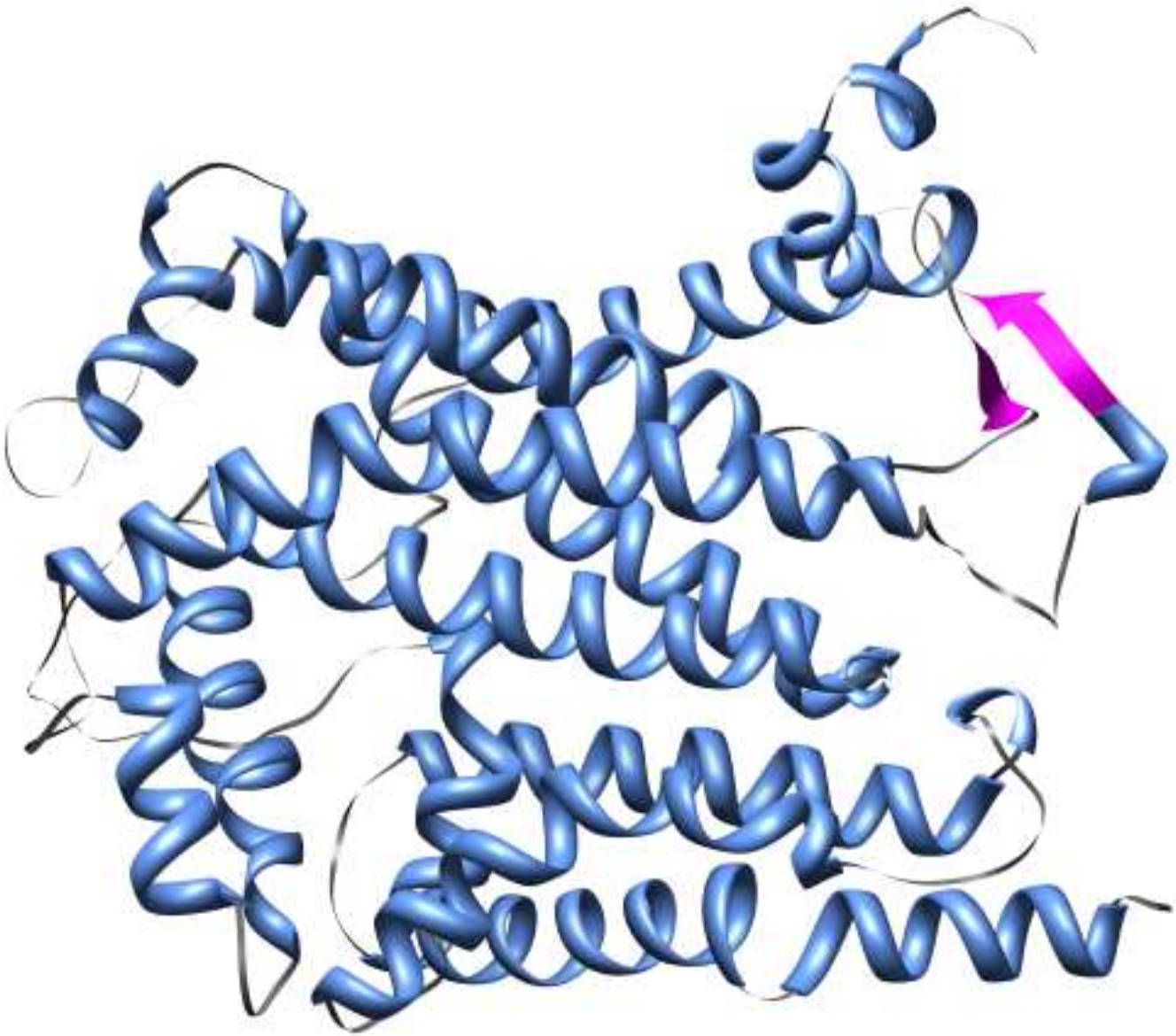
[1]Laboratory of Molecular Systematics, MSC, Smithsonian Institution, MRC-534, Washington, DC, 20560.
[2]To whom correspondence should be addressed at Department of Biological Sciences, University of Idaho, Moscow, Idaho 83844.

1064-7554/97/0600-0077$12.50/0

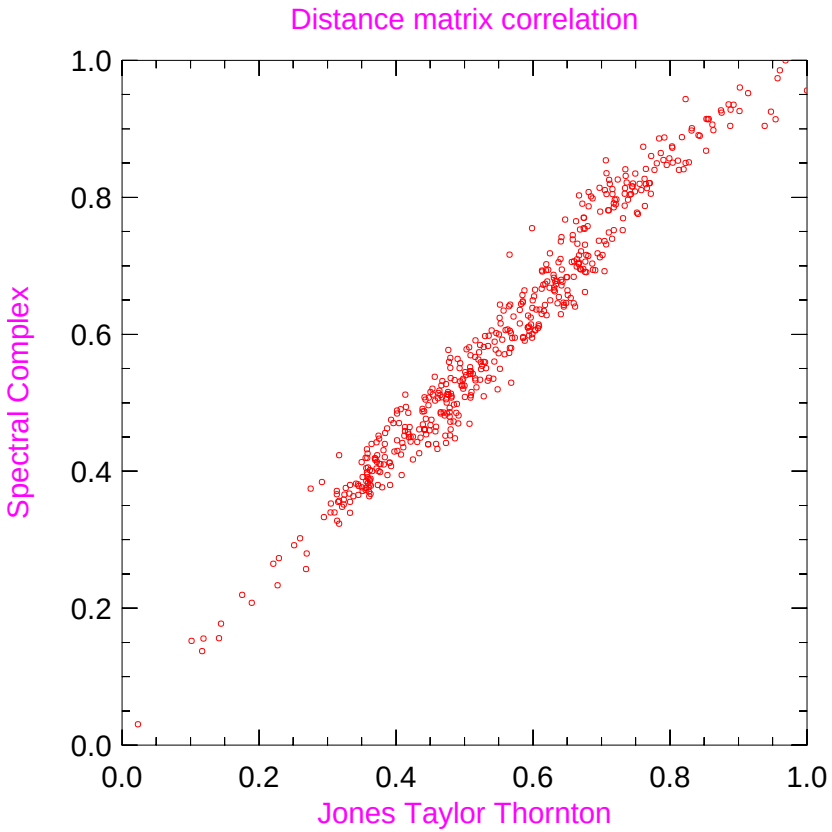
Challenging rodent data-set

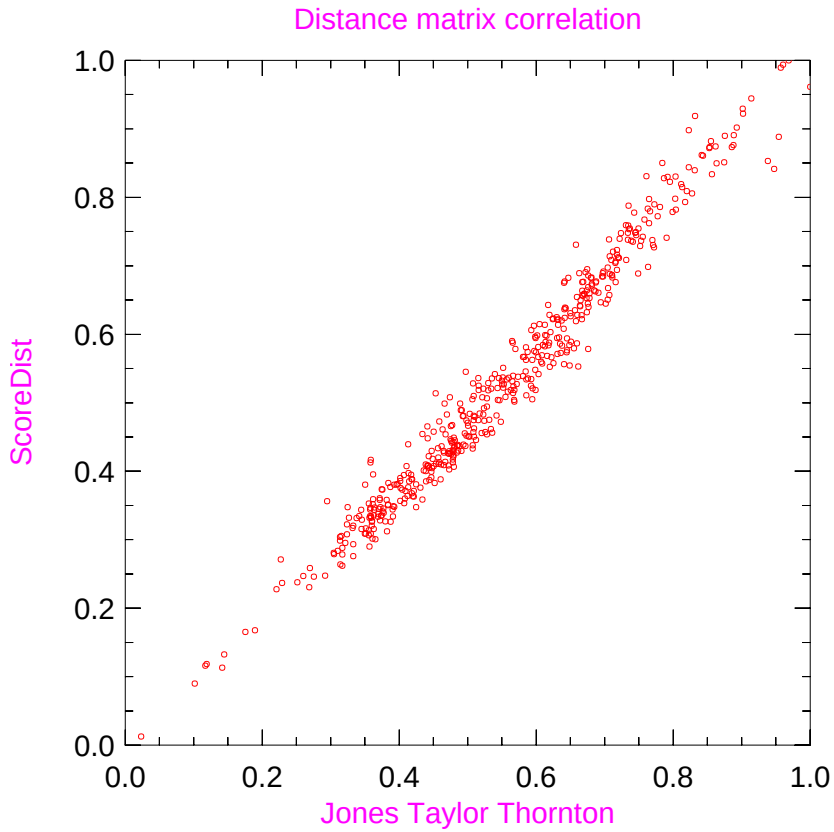
	Common name	**Scientific name**	**Short name**
Q36461	Platypus	Ornithorhynchus anatinus	Platypus
B2WVN5	Leopard	Panthera pardus	Leopard
O03207	White Rhinoceros	Ceratotherium simum	W Rhino
O47561	European Hare	Lepus europaeus	Hare
O63910	Dormouse	Myoxus glis	Dormouse
P00156	Human	Homo sapiens	Human
P00157	Cow	Bos taurus	Cow
P00158	Mouse	Mus musculus	Mouse
P00159	Rat	Rattus norvegicus	Rat
P24950	Fin Whale	Balaenoptera physalus	Fin Whale
P24959	Sheep	Ovis aries	Sheep
P24964	Pig	Sus scrofa	Pig
P34863	Rabbit	Oryctolagus cuniculus	Rabbit
P41285	Blue Whale	Balaenoptera musculus	Blue Whale
P41303	American Opossum	Didelphis marsupialis	A Opossum
P48665	Horse	Equus caballus	Horse

Molecular structure of Cytochrome-b

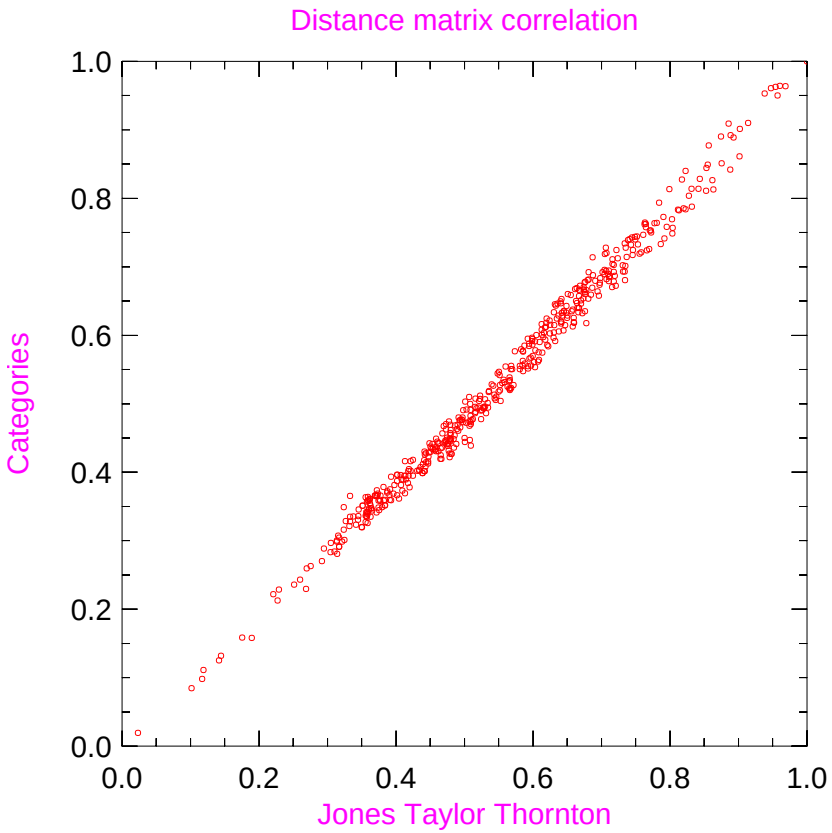


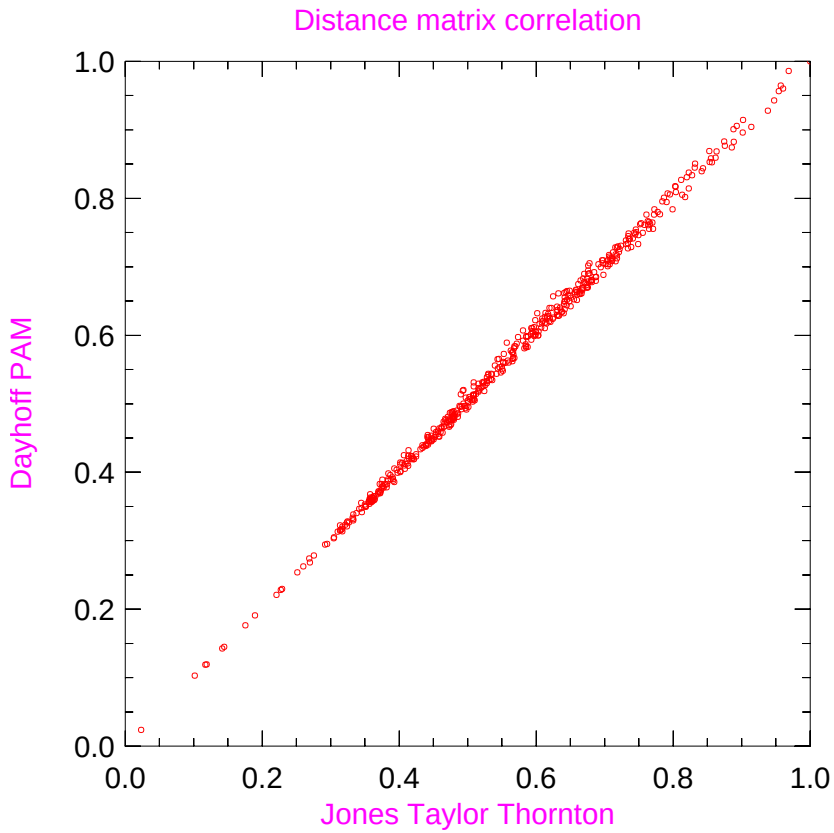
Comparison against established distance measures





More comparisons





"Goodness of fit" rankings

A remarkable result is that phylogenetic trees as determined by well established software packages have considerably lower least-squares residual values compared against other accepted distance measures. With respect to a phylogenetic tree T appropriate to the data, if $d_{j,k}$ is the sum of lengths (weights) for the edges that comprise the path between leaf j and leaf k and $D_{j,k}$ is the measures or estimated distance between the same leaves then the weighted sum of squared residual values is

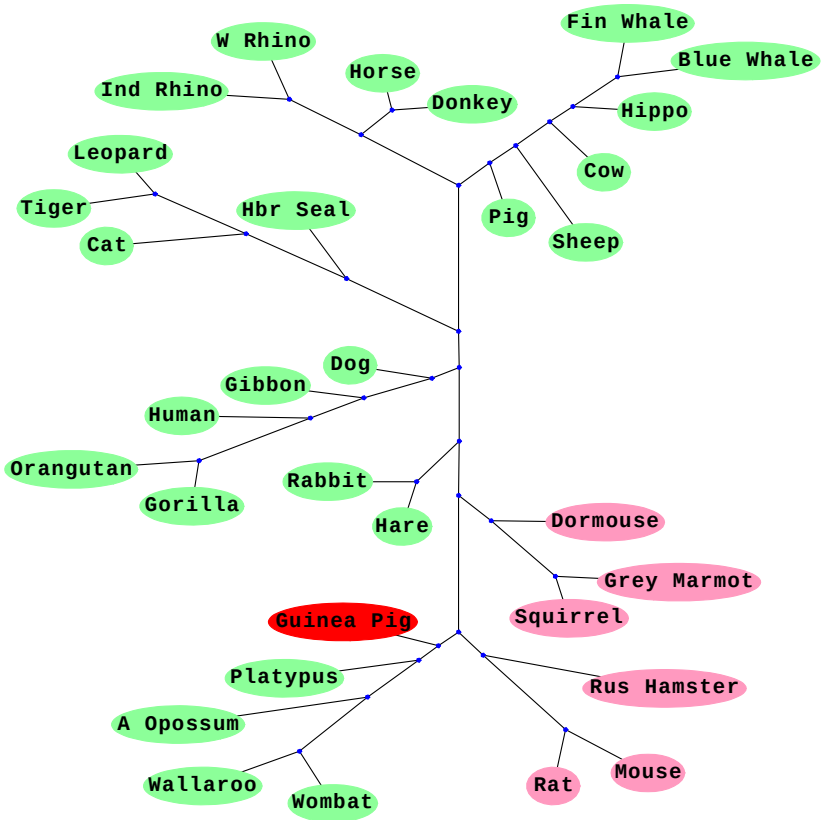
$$S(T) = \sum_{j,k\,\varepsilon\{L\}} \left(\frac{d_{j,k} - D_{j,k}}{D_{j,k}} \right)^2 ; \quad \{L\} \text{ is the set of all leaves}$$

Tree structure consistency

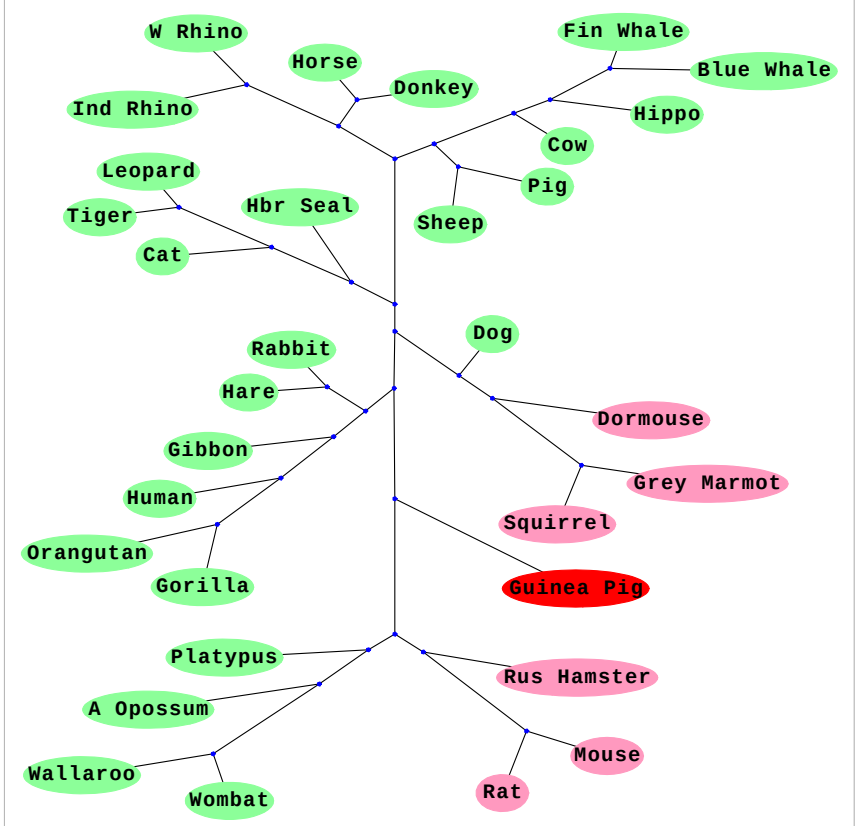
Distance measure\Tree algorithm	GNJ	Fitch-Margoliash	BIONJ
Spectral Complex	2.70953	2.75540	3.27449
Spectral Scalar	2.34364	2.34365	3.21843
ScoreDist	3.52057	3.52057	3.82674
Jones Taylor Thornton	3.60840	3.60840	4.00037
Henikoff/Tillier PMB	3.61953	3.62030	3.85857
Dayhoff PAM	3.71021	3.71021	3.99255
Kimura	3.85859	3.85859	4.24072
Categories	3.41611	3.41612	3.68582

Table: Weighted least-squares fit results returned from trees determined by the distance and tree-reconstruction algorithms listed, applied to the mitochondrial cytochrome-b sequences of the taxa from the previous data-set.

Cytochrome-b Phylogenetic trees I

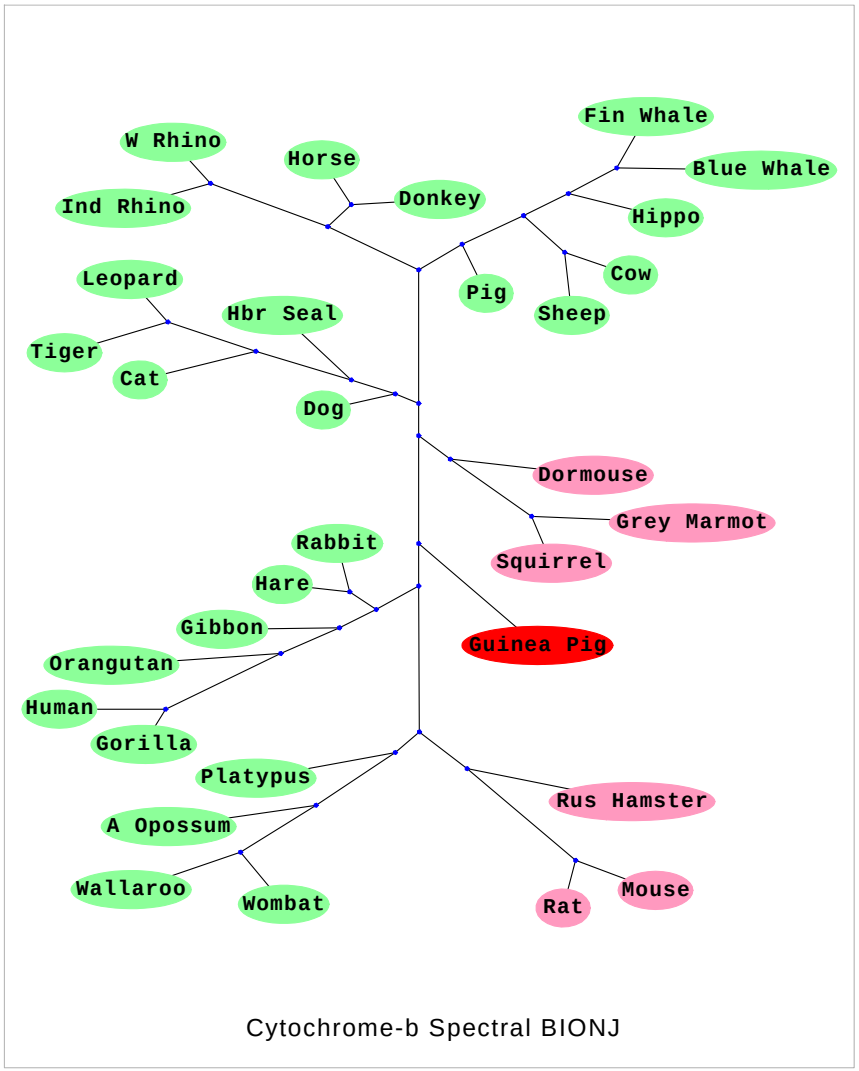


Cytochrome-b Categories BIONJ

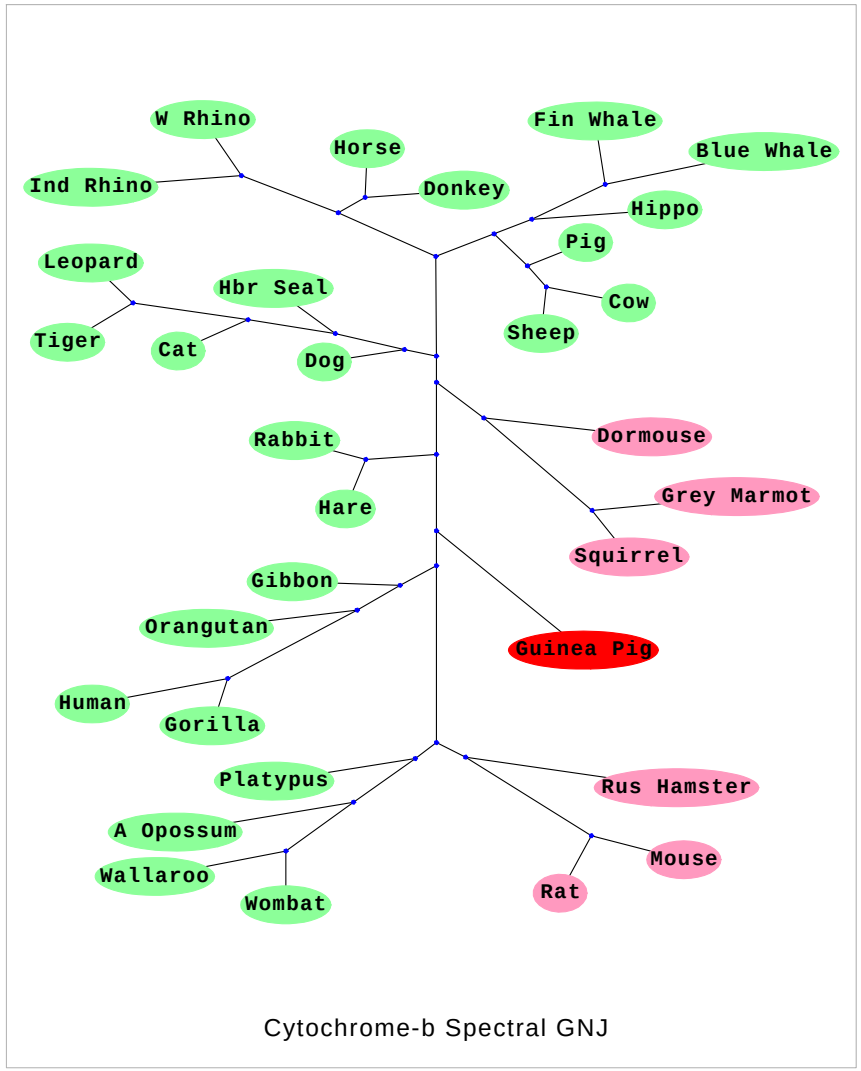


Cytochrome-b ScoreDist BIONJ

Cytochrome-b Phylogenetic trees II

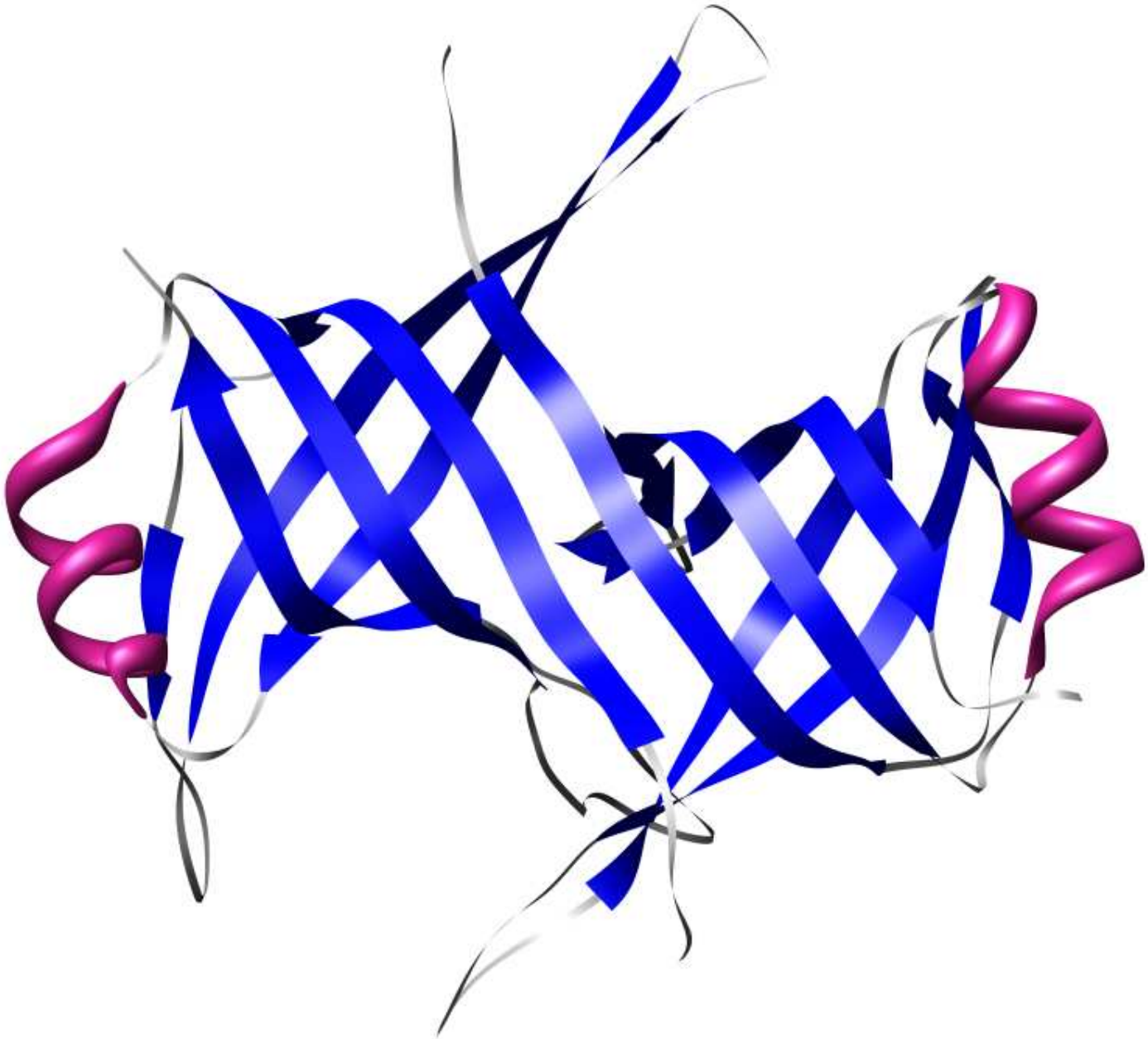


Cytochrome-b Spectral BIONJ

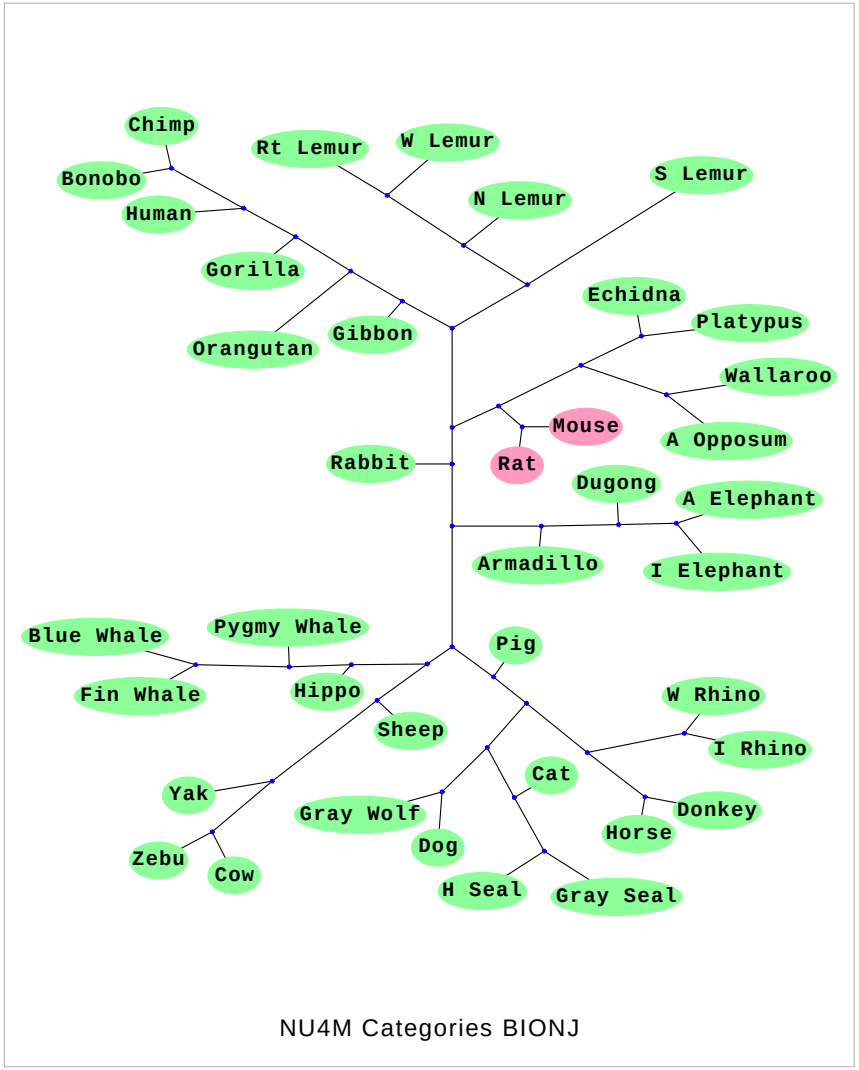


Cytochrome-b Spectral GNJ

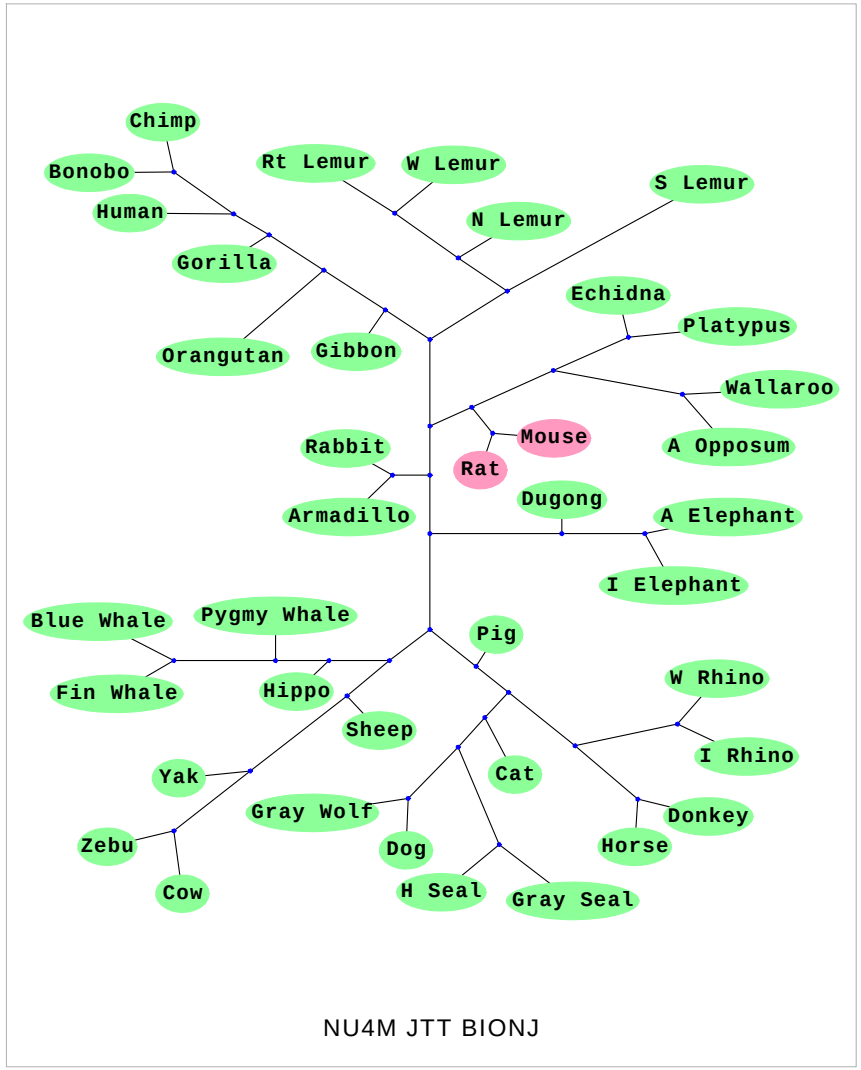
Two strands of Human NADH-ubiquinone oxidoreductase



NU4M Phylogenetic Trees I

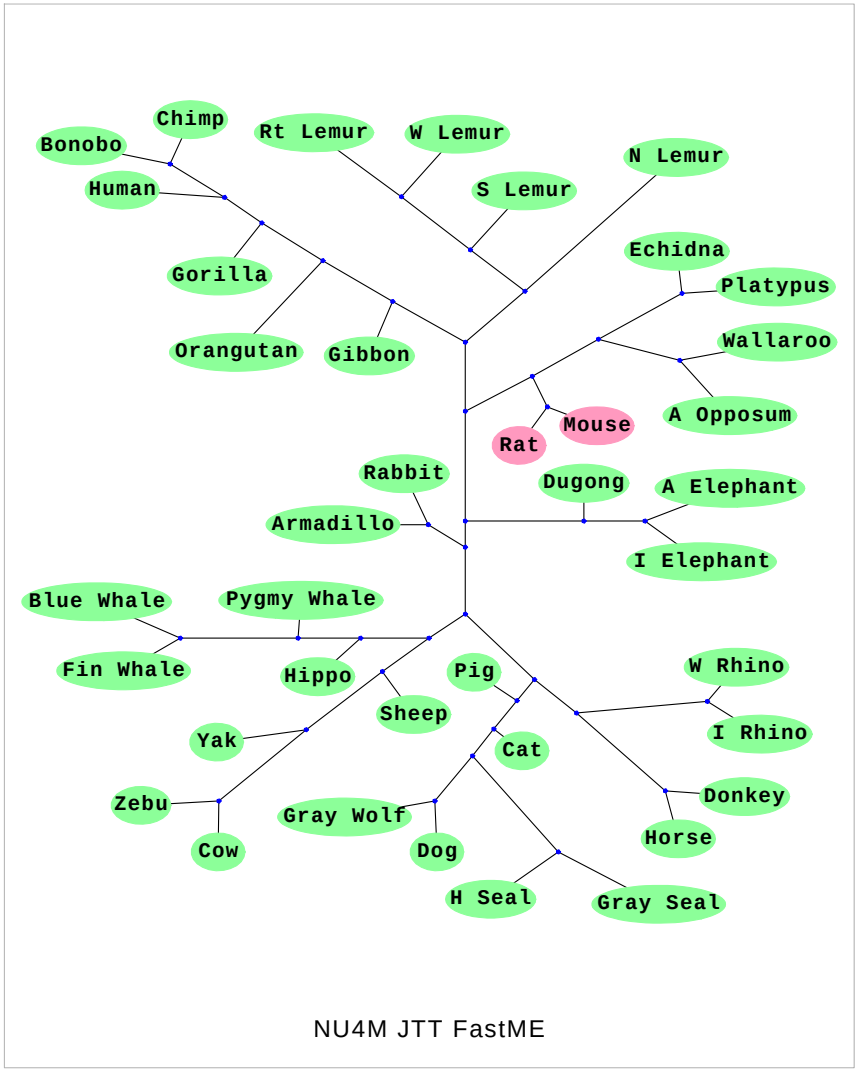


NU4M Categories BIONJ

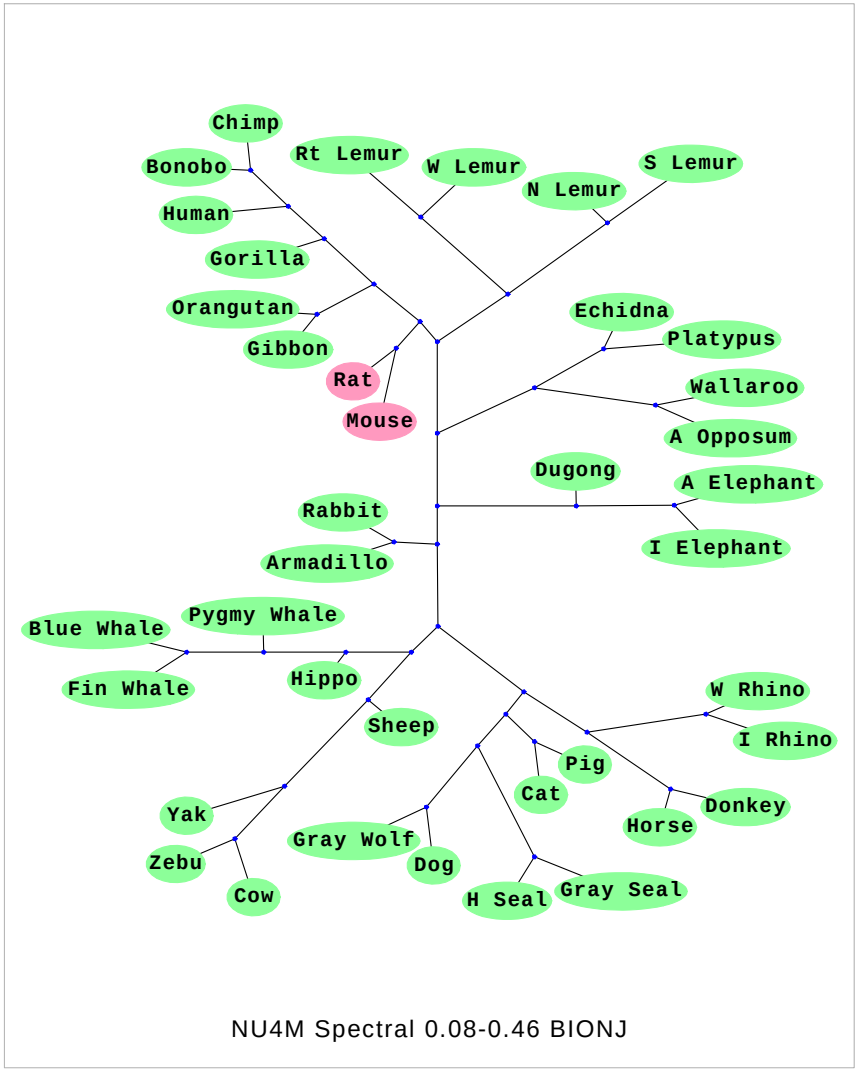


NU4M JTT BIONJ

NU4M Phylogenetic Trees II



NU4M JTT FastME



NU4M Spectral 0.08-0.46 BIONJ

Work in progress
(Second and third cytochrome-b α–helices)