

Do *Neruda* and *Laparus* stand outside *Heliconius*? JM 7 July 2006

The problem with the existing trees referring to this problem is that only *Heliconius* species have short branches. *Eueides* species, and the "basal" group of *Heliconius*, and all basal heliconiines tend to have long branches, and are apparently older than crown group *Heliconius*, such as the *melpomene-cydno-silvaniform* group, and to a lesser extent, the *erato* group (*erato-sara-charithonia* group tend to be rather distant species from one another).

In addition to problems of long branches in most of the basal groups (including *Neruda* and *Laparus*), the majority of the information so far brought to bear on the problem is based on either rapidly evolving, AT-rich mitochondrial data, which has a tendency to saturate very rapidly at the bases that do vary, and slowly evolving nuclear sequences such as EF-1a and wingless. For instance, of 3800 bp used by Beltrán et al. (submitted), 1600 or so was mtDNA, and the rest slowly evolving nuclear genes yielding little phylogenetic information. Thus the mtDNA dominates the information, even though saturation at its variable bases makes it unreliable.

Here I attempt to use intron sequences of Tpi, which seem to have a very rapid rate of evolution at all bases, to do the same trick.

Data sets

I used two data sets:

Tpi4.nex which had the original Tpi data from Margarita, plus *Neruda aoede* from Kanchon. Strangely, all three *Neruda* sequences were heterozygotes for the same indel (in such cases "--????--" has been put in to indicate the region of unreadable indel – Kanchon says it is only about 20bp in *Neruda*), and also heterozygotes for many of the single base differences. (Polyploid with two copies of the gene? Parthenogenetic? Strange. Anyway, to be explored).

The results are that in an alignment of 720 bp, *Neruda* (540 bp) and *Laparus* (590 bp) both have pretty good representation of the intron, including both *melpomene*- and *erato*-specific components. In contrast, the *erato* group (360-400 bp) tended to have a 3' intronic region with little overlap with the *melpomene* group (usually 510-530bp, though *hecale* and *numata* (560-610bp) had some of this "*erato*-specific" sequence), while the *melpomene* group tended to have a 5' region which no *erato* group species had. Thus both *Neruda* and *Laparus* retain some ancestral sequences that are deleted in both *melpomene* and *erato* sequences. However, the *Neruda* and *Laparus* sequences are highly divergent, both between the two genera, and also from the rest of *Heliconius*. Because *Neruda* and *Laparus* contain so much ancestral sequence, I am pretty confident that the alignments are mostly good approximations.

Tpi5.nex consisted of a smaller selection of the data from Tpi4.nex, selected to maximize diversity (including as many phylogenetic subgroups as possible), while at the same time maximizing sequence information (i.e. I focused on individuals that had good representation of intron sequence, and dropped individuals such as the *elevatus* which lacked the intron entirely – in retrospect I should have dropped *wallacei*, since it has very little sequence – 220 bp, and lacks almost all of the intron). I also attempted to root the tree using *Eueides isabella*, which was my most sequence-rich *Eueides*, the presumed closest relatives of *Heliconius*. Unfortunately, most of the

Eueides intron was unalignable. (*Eueides* species are alignable with each other, but only just. *Eueides* seems to consist of old species, with many differences – at mtDNA too). Essentially only the parts near the coding region could be aligned with other heliconiines. However, this does serve to give a baseline of divergence, with the interior of the intron being almost completely random. Although I was baffled trying to align the central region of the intron, I was able to align 185 bp of the total 369 bp of *isabella* sequence to *Heliconius* and *Laparus*.

Results

At Tpi, *Neruda* and *Laparus* are very divergent from *Heliconius*. This shows up in essentially all the trees which show branch lengths resulting from this data, with many parsimony informative sites. The divergence between each pair of *Heliconius*, *Laparus* and *Neruda* is high and approximately equal. *Laparus* can sometimes be slightly closer to some *melpomene* group species in some regions of the Tpi fragment, but this doesn't seem a very strong effect, because like a true ancestral species, it also has quite a few substitutions otherwise only found in *erato*.

Unless there has been a speed up of evolution in *Neruda* and *Laparus*, a molecular clock argument would suggest that both *Neruda* and *Laparus* stand outside of *Heliconius*, at least outside of the *Heliconius* for which we have Tpi data. We have essentially the full range of *erato*-group *Heliconius*, but we are lacking some of the more basal species such as *nattereri*, *egeria*, *hierax*, etc. from the *melpomene* group. But I don't suppose they will change things much.

Furthermore, rooting the tree with *Eueides* (on the grounds there's no doubt that *Neruda* and particularly *Laparus* are closer to *Heliconius* than to *Eueides*), leads to the same conclusion. I haven't systematically done bootstrap tests, but you can see in most MEGA bootstrap analyses that at least *aoede* is firmly distinct from *Heliconius*, even if you view the tree as unrooted.

In conclusion, Tpi seems to give pretty strong support for a traditional Brown phylogeny, with *Laparus* and *Neruda* basal to the rest of *Heliconius*. Judging by the divergences, there's a slight tendency for *Neruda* to branch off more ancestrally, with *Eueides* further out still.

Average HKY85 + gamma estimated distances (from Tpi5.nex), with alpha = 0.5. Expressed in raw fractions (i.e. 1.26 is 126%, 126 changes per 100 base pairs). Note distances between mel/cyd/silv to the era-sara-char group (green) are almost always less than *Laparus* or *Neruda* distances to *Heliconius* distances (blue)

Eueides							
Neruda	1.26						
Laparus	1.23	0.20					
silvaniforms	1.24	0.22	0.21				
melpomene-cydn	1.11	0.21	0.22	0.05			
erato	0.60	0.22	0.16	0.09	0.08		
sara-sapho	0.65	0.26	0.21	0.13	0.11	0.12	
ricini-charithonia	0.70	0.24	0.18	0.12	0.10	0.09	0.09
	Eueides	Neruda	Laparus	silvaniforms	melp-cyd	erato	sara

Overall data (here and below, missing pairwise comparisons are excluded)

Eueides							
Neruda	0.40						
Laparus	0.39	0.09					
silvaniforms	0.42	0.12	0.13				
melpomene-cydn	0.42	0.16	0.15	0.05			
erato	0.41	0.13	0.13	0.07	0.08		
sara-sapho	0.37	0.15	0.15	0.10	0.11	0.09	
ricini-charithonia	0.43	0.14	0.12	0.09	0.10	0.07	0.07
	Eueides	Neruda	Laparus	silvaniforms	melp-cyd	erato	sara

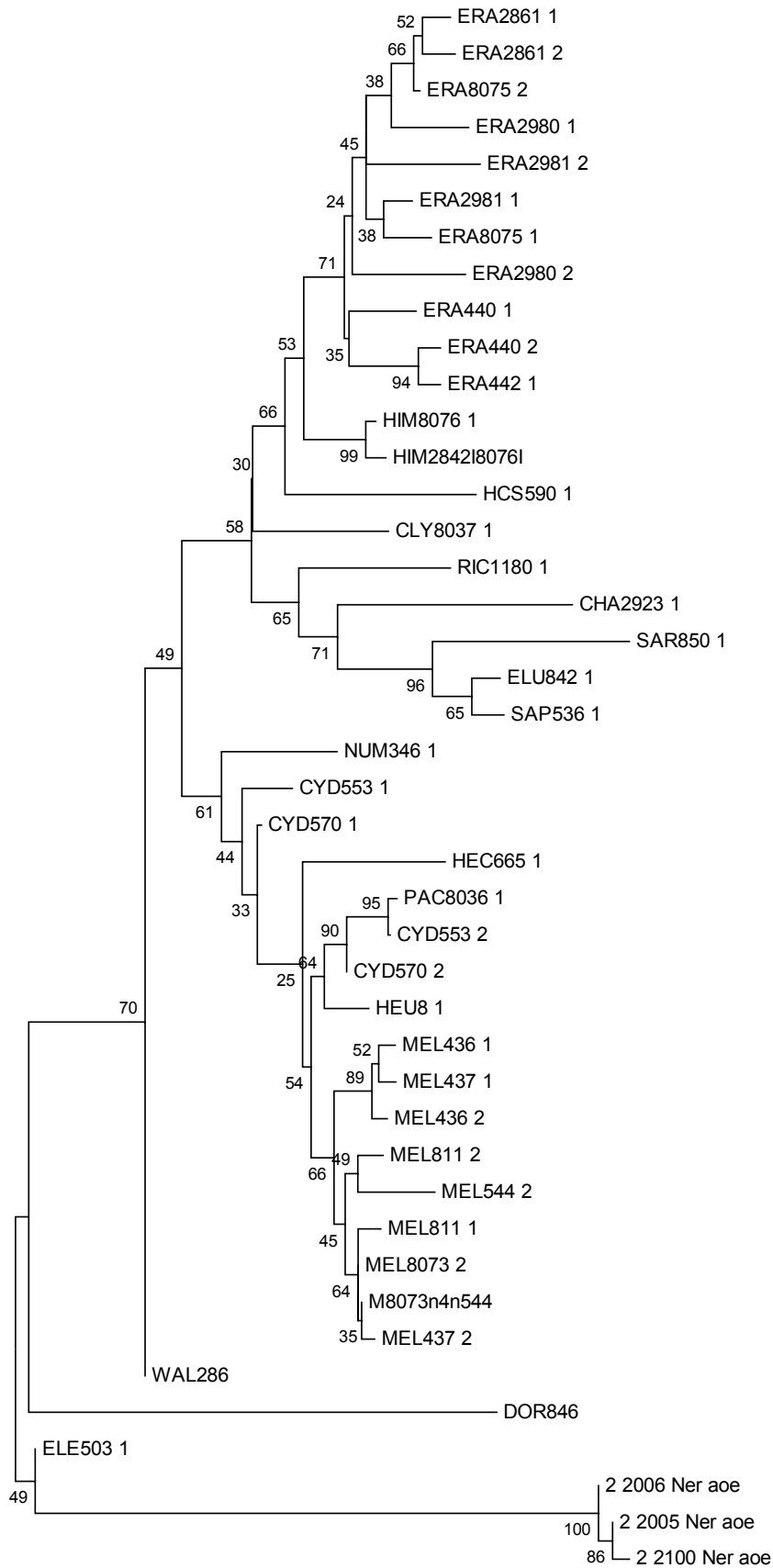
Sequence data from 5' and 3' coding + some adjacent intron overlap, and the central overlap of intron between *Neruda*, *Laparus*, *erato* and *melpomene* groups

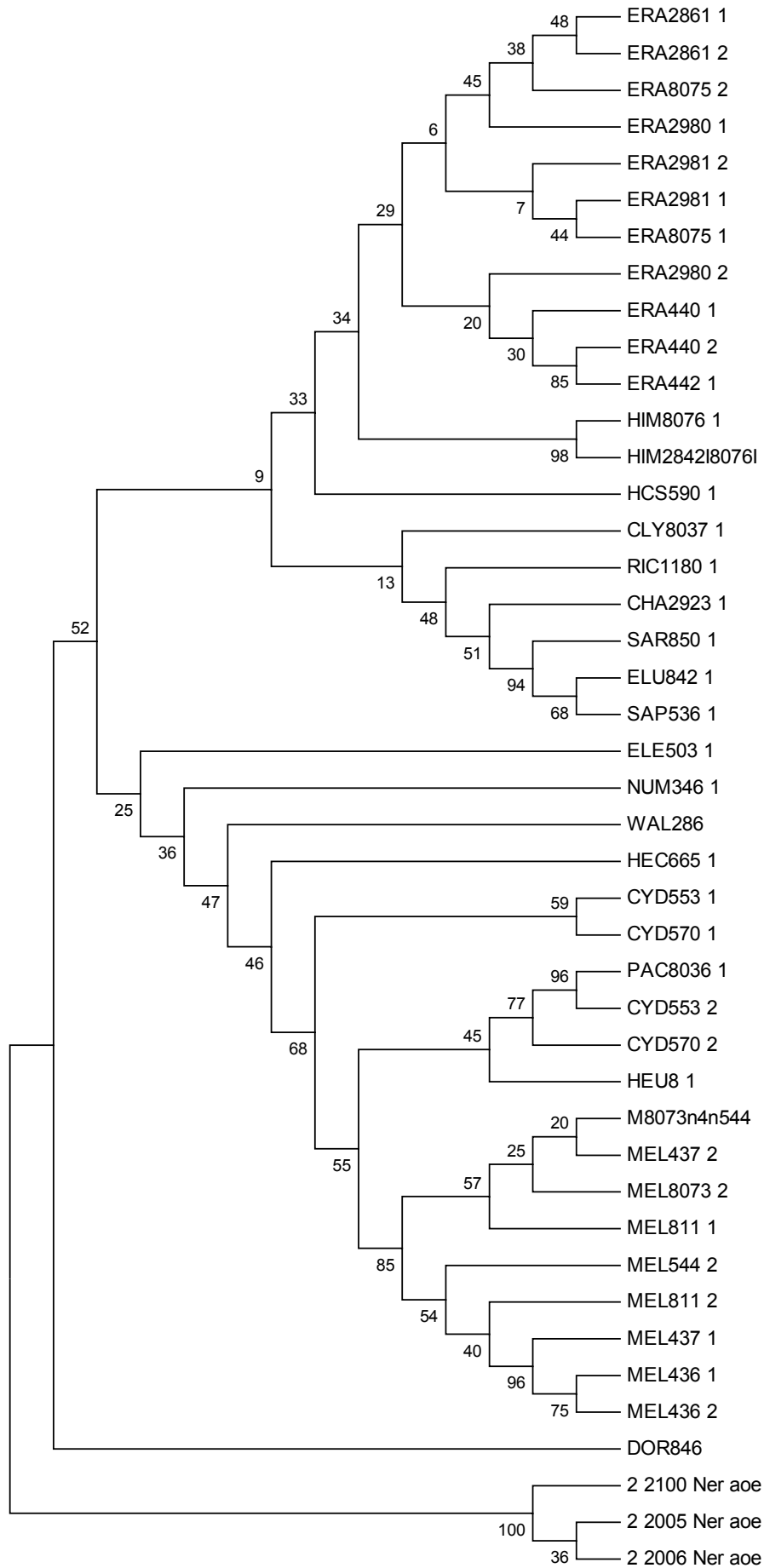
Eueides							
Neruda	*large						
Laparus	*large	0.74					
silvaniforms	*large	0.79	0.39				
melpomene-cydn	* large	0.15	0.68	*			
erato	* large	0.68	0.30	0.26	(0.02)		
sara-sapho	* large	0.87	0.47	0.35	(0.24)	0.22	
ricini-charithonia	* large	0.72	0.40	0.29	(0.11)	0.15	0.14
	Eueides	Neruda	Laparus	silvaniforms	melp-cyd	erato	sara

"*erato*-specific" mainly 5' regions of intron (note, the silvaniforms *hecale* and *numata* had substantial fractions of this sequence, but the *melpomene-cydn* group had very little, and divergences mel-cyd to the *erato* group are therefore not believable)

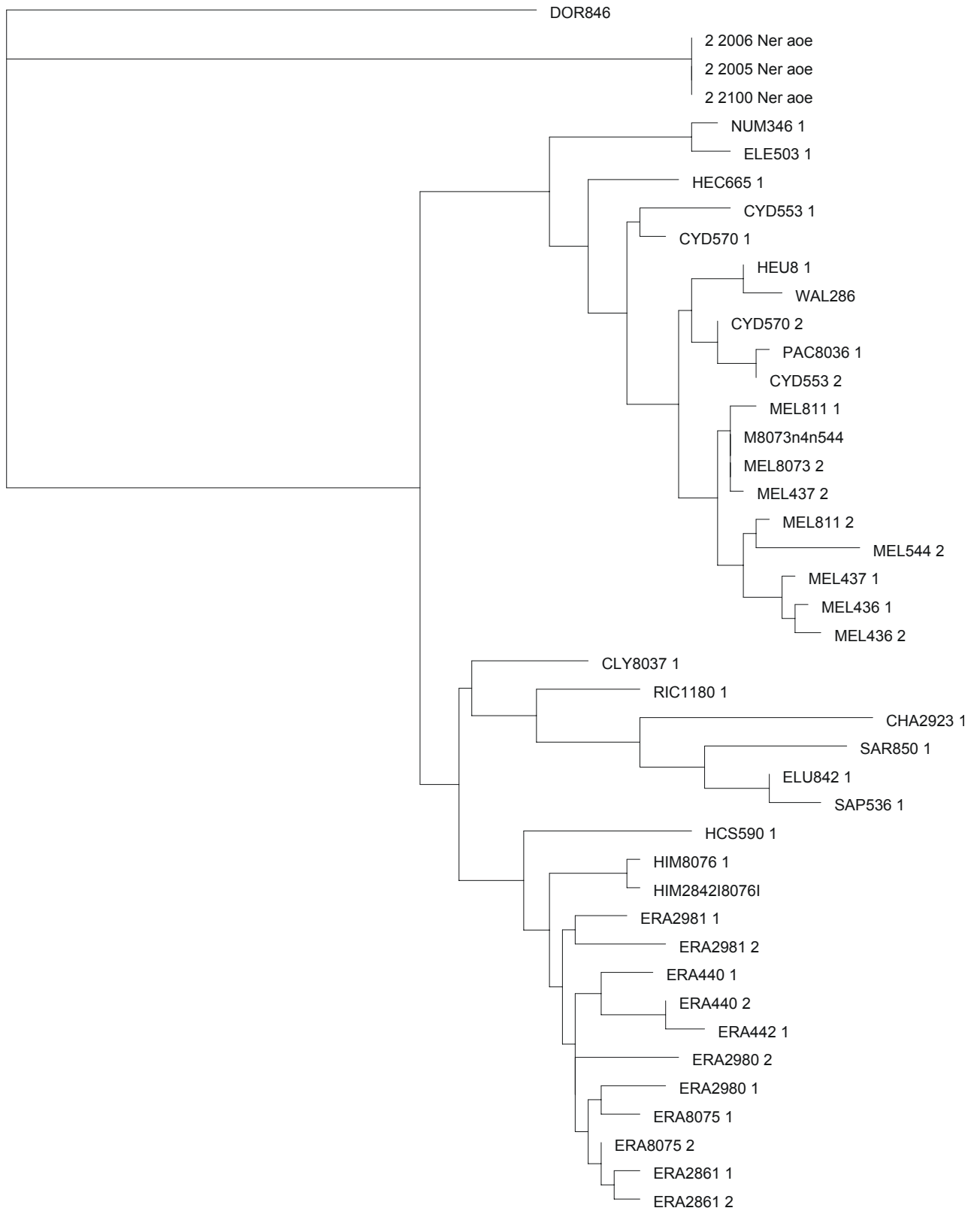
Eueides							
Neruda	* large						
Laparus	* large	0.19					
silvaniforms	* large	0.27	0.28				
melpomene-cydn	* large	0.32	0.31	0.05			
erato	* large	*	*	*	*		
sara-sapho	* large	*	*	*	*	*	
ricini-charithonia	* large	*	*	*	*	*	*
	Eueides	Neruda	Laparus	silvaniforms	melp-cyd	erato	sara

"*melpomene*-specific" mainly 3' regions of intron (no *erato* group species had this sequence region)



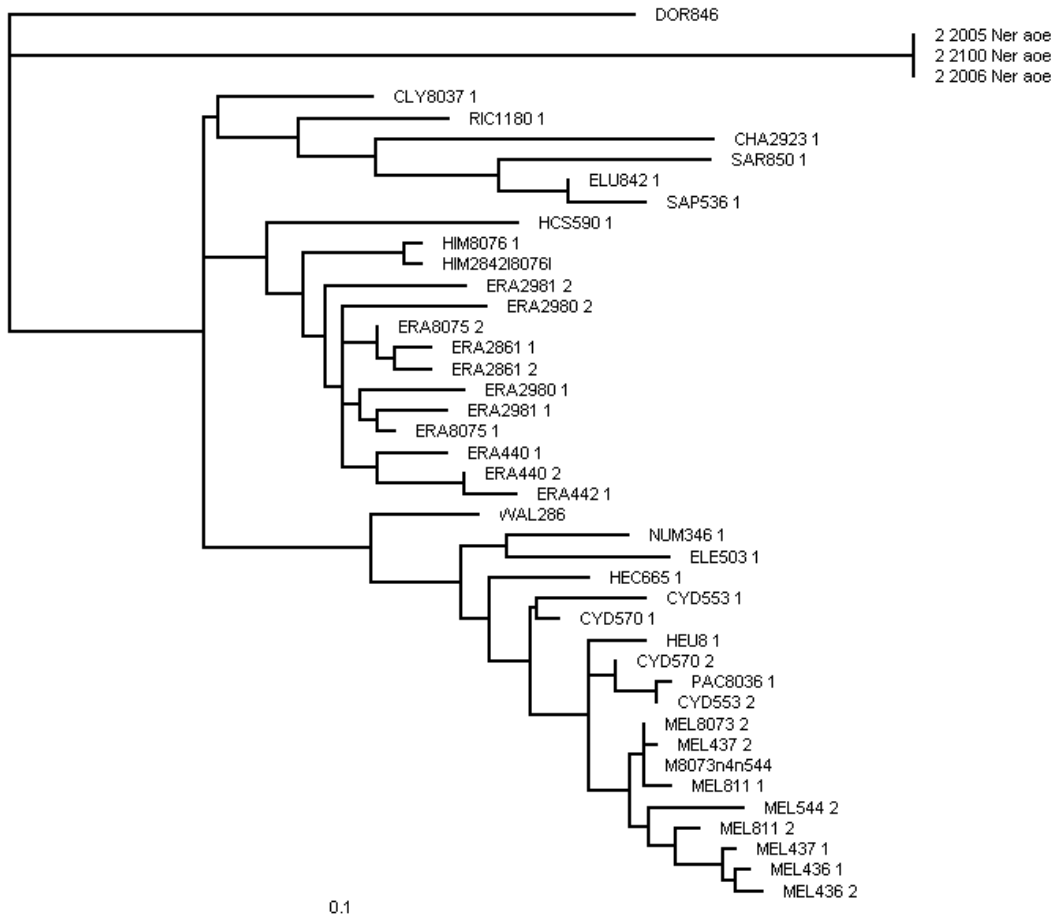


Parsimony Tpi4.nex tree, computed in MEGA (100 bootstraps)

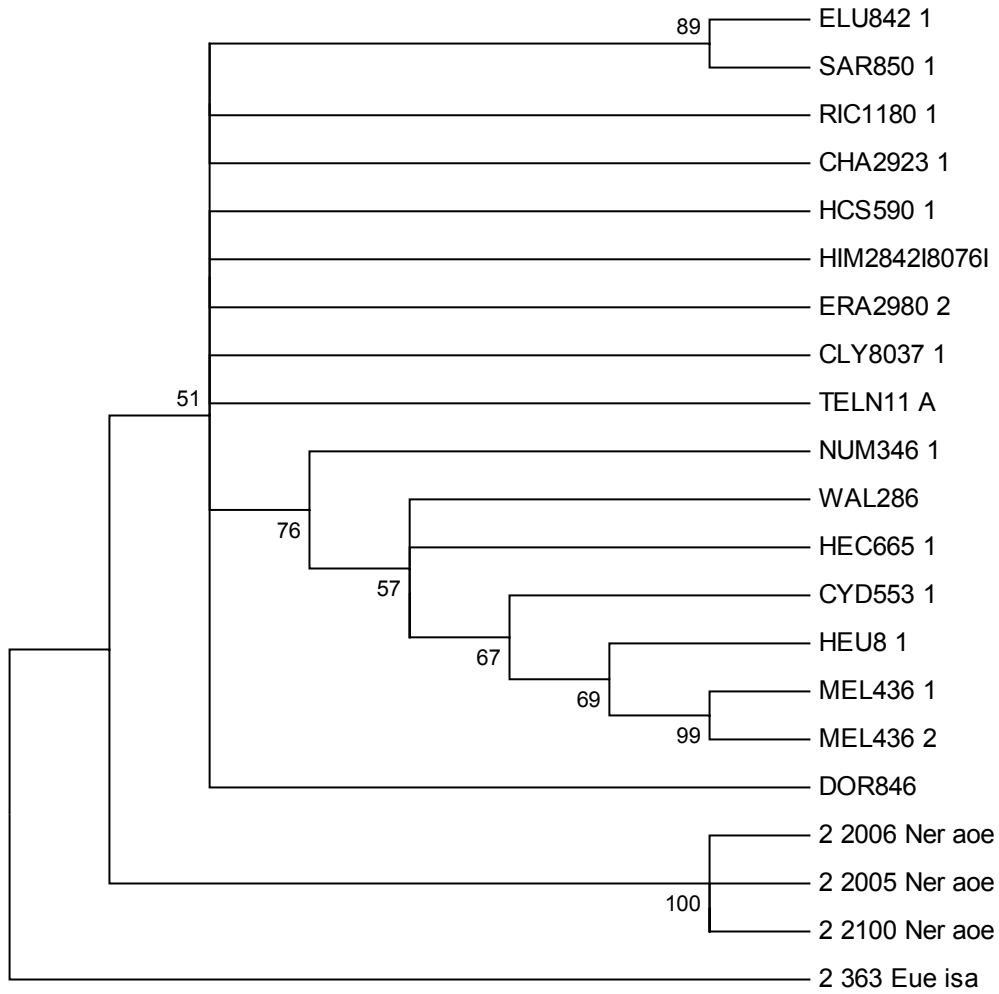


_1

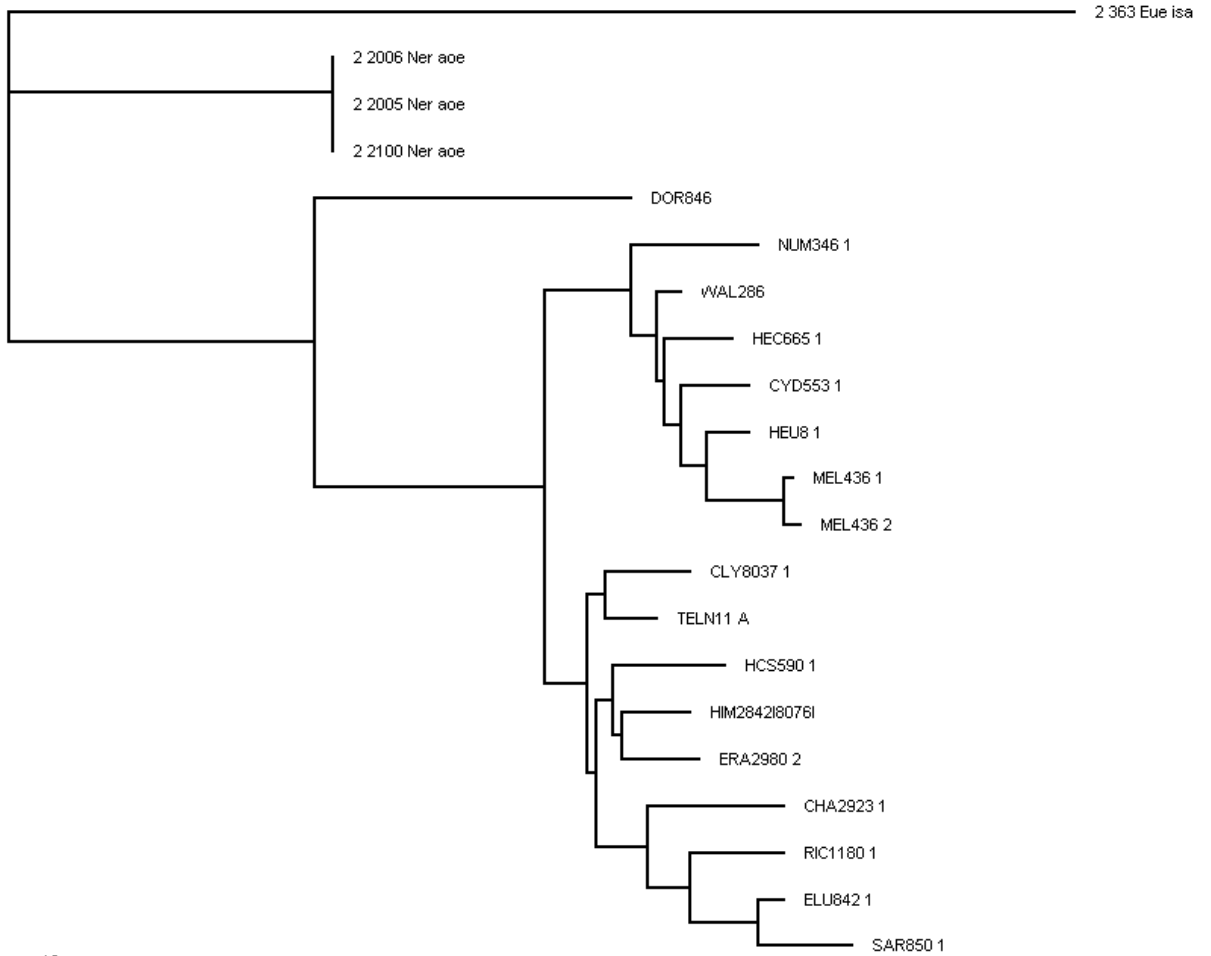
Parsimony Tpi4.nex tree, computed in MEGA (100 bootstraps)
With parsimony branch lengths added in PAUP



Tpi4.nex tree using PAUP ML and HKY85 + G rate variation.



Parsimony tree, reduced dataset Tpi5.nex (MEGA with 500 bootstrap values >50%)



Parsimony tree, reduced dataset Tpi5.nex (MEGA with PAUP parsimony branch lengths).