Available online at www.sciencedirect.com
ScienceDirect
Molecular Phylogenetics and Evolution 45 (2007) 757–762
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Short Communication

Evolutionary history of the genus *Sus* inferred from cytochrome b sequences

Stefano Mona [a,*], Ettore Randi [b], Mila Tommaseo-Ponzetta [c]

[a] *Department of Genetics and Microbiology, University of Bari, Italy*
[b] *Istituto Nazionale per la Fauna Selvatica, Ozzano Emilia (BO), Italy*
[c] *Department of Zoology, University of Bari, Italy*

Received 19 February 2007; revised 9 May 2007; accepted 30 May 2007
Available online 16 June 2007

1. Introduction

The systematic status of species belonging to the genus *Sus* has been a matter of debate for decades (Groves, 1981). According to a recent review of its taxonomy, there are a total of 7 living species (*Sus scrofa*, *Sus salvanius*, *Sus barbatus*, *Sus verrucosus*, *Sus celebensis*, *Sus philippensis*, *Sus cebifrons*) and approximately 22 subspecies (Groves and Grubb, 1993), to which one recently recognized species, *Sus bucculentus* (Groves et al., 1997) and many new subspecies (Groves, 1997) should be added.

South-East Asia (SEA) may be considered the homeland of the genus, as 6 out of the 8 species are endemic to the area, the only exception being the pygmy hog (*S. salvanius*), nowadays present only in the north of India, and *S. scrofa*, which is distributed worldwide (and not limited to SEA).

In particular, the relationship between SEA species was investigated, but the evidence presented until now is still not conclusive. Morphological analyses point toward the identification of two main lineages, one leading to *scrofa/celebensis/philippensis* and the other to *barbatus/verrucosus* (Groves, 1997), but molecular evidence does not support the same grouping, but merges *scrofa*, *barbatus* and *verrucosus* (Lucchini et al., 2005). Recent findings have further complicated this issue rather than solving it, highlighting a substantial level of polyphyly among some of these species (see Larson et al., 2005).

There is also great uncertainty about the temporal scale of *Sus* evolution, mainly because of a general lack of representative fossil records. For example, the origin of the genus probably dates back to near the Miocene/Pliocene boundary, around 5 Mya[1] (Randi et al., 1996), but its divergence from its closest genus, *Phacochoerus*, based only on molecular data, ranges from 5 to 15 Mya (Randi et al., 1996). This great uncertainty also characterizes estimates about the separation between European and Asian lineages of *S. scrofa*, spanning from 56,000 (Kim et al., 2002) to 500,000 years (Giuffra et al., 2000), depending on the marker and mutation rate used.

The major aim of the present study was to provide a more comprehensive description of the timing of evolution of the genus *Sus* and of the phylogenetic relationships among its various species. We thus sequenced mitochondrial cytochrome b from some specimens belonging to *S. scrofa*, *S. celebensis* and *S. barbatus*. This gene was selected as it was the most represented marker available for *Sus* in the GenBank. We further tested the performance of the cytochrome b as a molecular barcode, in order to check its ability to recognize potential hybrid populations such as that inhabiting New Guinea, proposed to be a cross between *S. celebensis* and *S. Scrofa* (Groves, 1981; but see Larson et al., 2005, for a different opinion).

2. Materials and methods

We collected 112 hair samples of *S. scrofa* from all over the world, 5 *S. barbatus* from Borneo and 3 *S. celebensis*

* Corresponding author. Present address: Department of Biology and Evolution, University of Ferrara, Via L. Borsari 46, 44100, Ferrara, Italy. Fax: +390532249761.
E-mail address: stefano.mona@unife.it (S. Mona).

[1] *Abbreviations used:* B.P., before present; Mya, millions of years; BF, Bayes factor; TMRCA, time to the most recent common ancestor.

1055-7903/$ - see front matter
doi:10.1016/j.ympev.2007.05.025

from Sulawesi. The *S. scrofa* samples have been already typed for the D-loop and encompass much of the mitochondrial variability of the species (Mona et al., unpublished). Sequence data have been deposited in GenBank under Accession Nos. AM492546 to AM492665. Additional sequences from *S. scrofa*, *S. celebensis*, *S. barbatus*, *S. verrucosus*, *S. bucculentus* and *S. philippensis* were extracted from GenBank, together with that of *Phacochoerus africanus* (common warthog). A complete and detailed list of samples is given in Table S1 of the Supplementary materials, and the geographic distributions of the species is shown in Fig. 1.

Total DNA was extracted following Gerloff et al. (1995). We decided to type cytochrome b since it provided the largest dataset for *Sus* in the GenBank. The marker was typed following the procedure of Alves et al. (2003). Electropherograms were checked by EDITVIEW (ABI PRISM, version 1.0.1) and the final sequences were aligned by eye with GENEDOC (http://www.psc.edu/biomed/genedoc).

Model selection of nucletotide substitution was performed with MRMODELTEST (Nylander, 2004) according to the AIC criterion. Phylogenetic inferences were performed with MRBAYES 3.1.2 (Ronquist and Huelsenbeck, 2003): each analysis was the result of two runs of 4 incrementally heated chains with default temperature, 10^7 generations long with a 10% burnin. Convergence was checked by examining the generation plot visualized with TRACER (Rambaut and Drummond, 2004) and computing the potential scale reduction factor (Gelman and Rubin, 1992) with *sump* in MRBAYES. Molecular dating, both enforcing and relaxing the clock, were achieved with BEAST (Drummond et al., 2006) with analogous settings used for MRBAYES runs (except for heating); output was summarized using *sump* and *sumt*. Within the Bayesian framework, model choice and hypothesis testing in phylogenetics is usually achieved through the Bayes factor (Nylander et al., 2004; and many others). However, the error in their estimations may be quite large (Suchard et al., 2005), and their dependence on model priors (Kass and Raftery, 1995) is not well-known in the phylogenetic context. Thus, together with the Bayes factor, we decided to assess model performance using a posterior predictive test (Gelman et al., 1996) following the approach of Bollback (2002). As discrepancy variable we chose multinomial test statistics (Goldman, 1993) as a general measure of model fit. Posterior predictive distribution was evaluated through Monte Carlo simulations of 1,000 datasets for each model using posterior densities of model parameters (tree topology, branch lengths and substitution parameters) inferred by MRBAYES or BEAST; MAPPS software (Bollback, 2002) was used for simulations. The discrepancy between observed test statistics and simulated predictive distributions in the various models was quantified using Bayesian *p*-values (Gelman et al., 1996) and the L-criterion proposed by Laud and Ibrahim (1995), both computed with MAPPS. The *p*-value was estimated as the proportion of the total number of Monte Carlo simulations for which the multinomial statistics was equal to or exceeded its realized value. The L-criterion takes into account the (square) deviation of the observed value of the test statistics from the mean of the predictive distribution, and adds a penalty due to the variance of the distribution: the lower the L-criterion, the better the model fits the data.

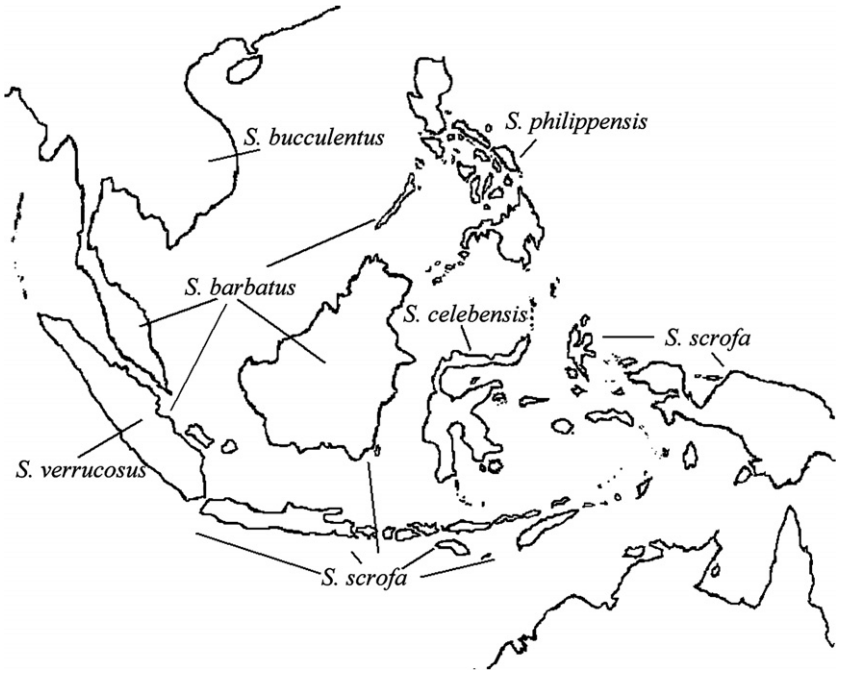

Fig. 1. Broad geographic distribution in South-East Asia of various *Sus* species examined in present study.

Population genetics parameter were estimated with BEAST with two runs of 2×10^7 generations with a 10% burnin using the nucleotide substitution model selected by MRMODELTEST and a coalescent prior with exponential growth. Distance analyses to determine species attribution were performed with TREEPUZZLE 5.2 (Schmidt et al., 2002) under different models of nucleotide substitution.

3. Results and discussion

3.1. Phylogenetic inference and molecular dating

In total, we obtained 172 complete cytochrome b sequences from six *Sus* species. Haplotypes were found with CLEANUP software (Grillo et al., 1996) and phylogenetic analyses were then performed on 66 *Sus* sequences and one *Phachocoerus africanus* as outgroup. Of the 24 nucleotide substitution models tested by MRMODELTEST, the GTR+Γ gave the lowest score under the AIC criterion, and was thus used for the following Bayesian and maximum likelihood tests. Since it is well-known that variation in rates of evolution along different lineages can severely affect phylogenetic estimation (Ho and Jermiin, 2004), we inferred relationships among *Sus* species using the standard unrooted approach (every branch with its own rate of evolution), the strict molecular clock (the same rate for all branches), and the two relaxed clock phylogenetic models recently proposed by Drummond et al. (2006), in which rates of evolution are uncorrelated among branches but all belong to the same distribution (modeled as exponential or lognormal). We employed a pure birth model (birth rate being a hyper parameter with a uniform prior) and a uniform model (branch lengths with a uniform prior) to detect the possible effect of prior assumption on topology and branch lengths (Welch et al., 2005; Yang and Rannala, 2005) in the strict clock and the two uncorrelated models. The latter failed to identify the root correctly (that is, the warthog outgroup) under uniform priors, and were thus discarded. Performance in terms of the Bayes factor and posterior predictive test of the remaining five models is listed in Table 1. Tree topologies visited by the Monte Carlo Markov Chain were summarized, for each model, with the 50% majority consensus rule; the five models resulted in two different consensus topologies, shown in Fig. 2. Topology *a* was derived from both the two strict clock (CU and CP) and the uncorrelated relaxed clock lognormal model (LOG). Topology *b* was derived from both the unrooted (UF) and uncorrelated relaxed clock exponential model (EXP). The Bayes factor and posterior predictive test led to radically different conclusions as regards the model to be chosen. In addition, the influence of the priors on the harmonic mean of the marginal likelihood (and thus on Bayes factor estimation) was remarkable. Indeed, when a pure birth prior was employed, following Kass and Raftery's (1995) scale for the Bayes factor, the clock model was slightly favoured compared with the unrooted one (BF = 6.22, computed as twice the difference in the log likelihood). Conversely, when a uniform prior was employed, the unrooted model was strongly preferred to the clock one (BF = 56). However, the predictive test showed that the multinomial statistics observed in our data almost fell out of the range of that simulated with the clock model with pure birth prior (Table 1), suggesting that this model is inadequate. Instead, the clock model with uniform prior was the best predictor of our data (Table 1). The reason for this discrepancy was probably due to the difference in the estimates of total tree length and root length of the two strict clock models, which had a profound effect on the simulated multinomial test statistic. It should be noted that the uncorrelated models also failed to reproduce the data adequately, although they would have been preferred by a Bayes factor comparison with strict clock and unrooted models. We thus suggest that a posterior check should always be performed, in view of its more direct and clear interpretation compared with the Bayes factor, and several discrepancy variables should be developed to capture specific features of the data. The effects of informative priors also need to be accounted for in any Bayesian analysis.

Table 1
Phylogenetic models used and their performance

Model	L[b]	E[Ts][c]	*p*-value[d]	Harmonic mean[e]
UF[a]	20243	−2780.76	0.72	−3606.44
CU	19822	−2790.97	0.67	−3634.48
CP	40119	−2689.01	0.91	−3603.33
LOG	56989	−2654.74	0.95	−3588.14
EXP	45201	−2676.94	0.93	−3573.39

[a] Models used: UF, unrooted model; CU, strict clock with uniform priors; CP, strict clock with pure birth priors; LOG, uncorrelated relaxed clock lognormal model with pure birth priors; EXP, uncorrelated relaxed clock exponential model with pure birth priors. See text for details.
[b] L-criterion of Laud and Ibrahim (1995). See text for explanation.
[c] Expected value of the multinomial test statistics under the corresponding model. The value observed in the data was −2850.11.
[d] Bayesian *p*-value of the observed test statistics.
[e] Estimated harmonic mean of the marginal likelihood.

In summary, using the posterior predictive test, CU was found to perform better than any of the other models (Table 1), as it gave the lowest L-criterion score and the best *p*-value. This implies that cytochrome b sequences of the genus *Sus* evolved in a very clocklike fashion. To further support this finding, we performed an LRT test on topology *a*, and the molecular clock, as expected, was not rejected (*p*-value = 0.12).

Interestingly, assuming that the origin (species) of DNA samples have been accurately identified, and that species distinction is correct, non-monophyly is quite widespread among species belonging to the genus *Sus*, since there are examples of both paraphyly and polyphyly. Taking topology *a* of Fig. 2 as our best estimates, we observed several cases of non-monophyly. First, *S. barbatus* is paraphyletic, due to the *S. verrucosus* being nested within the *S. barbatus* haplotypes: this relationship is strongly supported (posterior probability of 1.00). Second, the posterior probablility

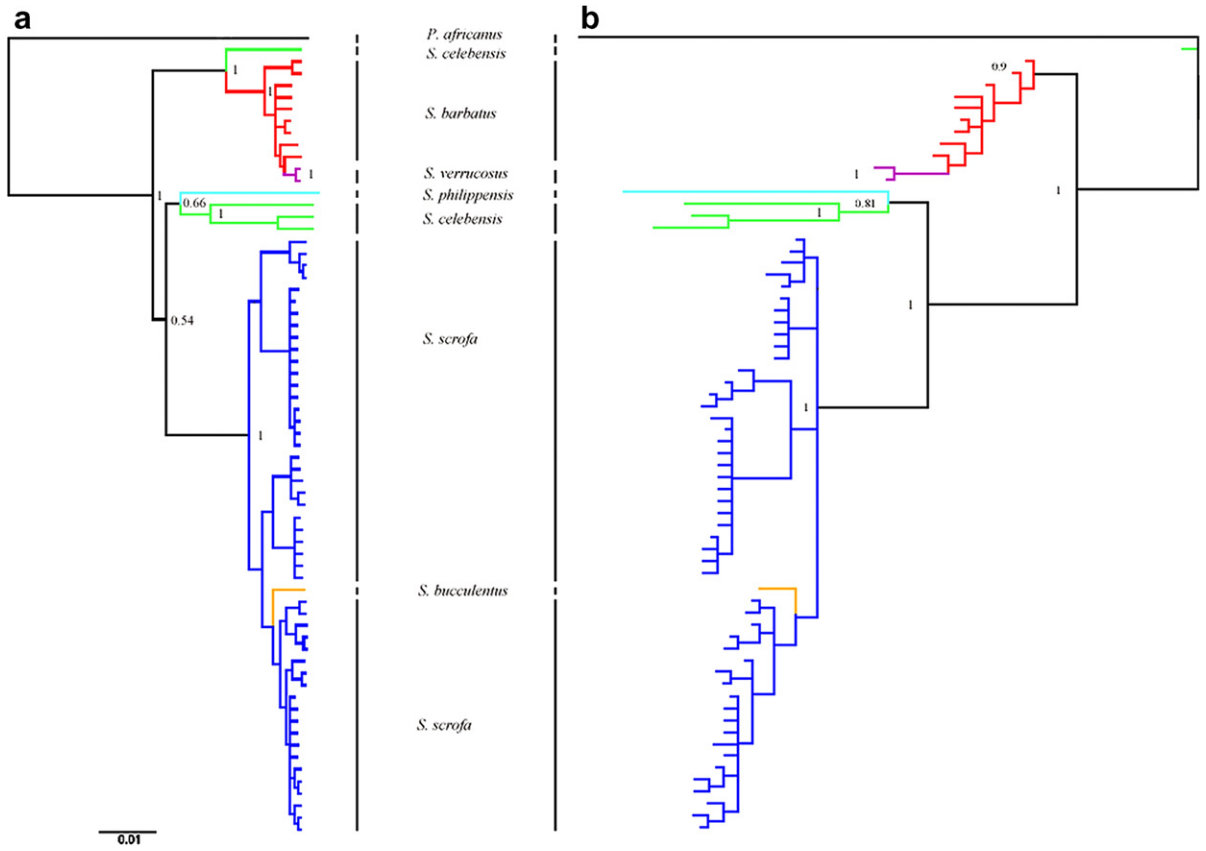

Fig. 2. (a) 50% majority consensus rule tree, based on cytochrome b sequences analyzed under strict clock model with uniform priors. Posterior probabilities of nodes relevant to species relationships are shown. (b) 50% majority consensus rule tree, based on cytochrome b sequences analyzed under unrooted model.

of *S. celebensis* haplotypes being monophyletic is zero (therefore *S. celebensis* is polyphyletic). Finally, *S. scrofa* is paraphyletic given the position of *S. bucculentus* (posterior probability of 1.00); however, as pointed out by Robins et al. (2006), *S. bucculentus* could be a geographically restricted variant of *S. scrofa* rather than a distinct species.

To further test this result, we performed two more Bayesian searches (settings as in CU). The first constrained all these species as monophyletic; the second only constrained *S. celebensis* monophyly, given the question of the status of *S. bucculentus* (see Robins et al., 2006). Since we specified the same model priors, we could then evaluate the Bayes factor in favor of the unconstrained clock model, and obtained estimates of 97.3 and 76.5, respectively; according to the scale of Kass and Raftery (1995), this may be taken as definitive evidence in favor of the unconstrained model.

In summary, given the high level of polyphyly and the low posterior probabilities of various clusters of species, we find no support either for the results of Lucchini et al. (2005) or for those of Groves (1997): the branching pattern of *Sus* phylogeny is far from being deciphered, as cytochrome b alone is not sufficient to resolve multifurcation. Clearly, an evolutionary radiation of the genus occurred indicating that only more markers (specifically, Y chromosome and autosomal loci) could be helpful in uncovering the relationship among the various species.

To frame the evolution of the genus *Sus* in a time-scale contest, we performed molecular dating. The molecular clock was calibrated placing the arrival of *S. scrofa* in New Guinea at 8,000 years B.P. (Mona et al., unpublished): 32 sequences of New Guinea pigs were used to estimate their TMRCA in units of coalescence time (New Guinea sequences belonging to cluster D6 following Larson et al. terminology; Table S1), and this was translated into a mutation rate using the above calibration point. The resulting posterior distribution of the mutation rate was then approximated by a normal N $(2.37\times10^{-8}; 6.05\times10^{-9})$ and used as a prior in the strict clock and uncorrelated relaxed clock model analyses. The results are shown in Table 2. The main difference among the four models lay in the estimate of the divergence time between *Sus* and *Phacochoerus*: the two strict clock models showed a substantial overlap of their 95% high posterior density, whereas the two uncorrelated models were almost out of the strict clock range and had a standard deviation of one order of magnitude larger. This was not observed for the coalescent time of the various species, whose estimates by the several methods were quite concordant (Table 2). However, as stated above, only the strict clock model with uniform prior was a good predictor of the data observed, and thus only its estimates are the reliable ones. Thus, the divergence between *Sus* and *Phacochoerus* is younger

Table 2
Coalescent time estimates

CLADE	CU[a]	CP[a]	LOG[a]	EXP[a]
Root[b]	2,407,585[d] (1,384,554–4,464,078)	3,451,927 (2,058,332–6,075,084)	8,980,618 (2,213,024 - 32,842,727)	17,463,844 (2,259,285–77,984,492)
Sus[c]	1,269,712 (734,328–2,340,764)	1,104,297 (659,832–1,964,261)	1,004,334 (497,207–2,127,404)	1,133,349 (523,054–2,326,629)
Sus scrofa[c]	459,173 (253,299–859,308)	358,132 (204,513–646,998)	401,002 (209,280–826,909)	413,892 (218,467–838,054)
Sus barbatus[c]	292,989 (145,711–581,663)	228,921 (120,481–435,354)	294,001 (127,484–678,172)	362,510 (149,183–825,958)
Sus verrucosus[c]	32,284 (5985–87,209)	24,957 (4759–66,294)	29,622 (4911–89,220)	31,503 (5171–91,665)
Sus celebensis[c]	1,204,527 (689,074–2,205,366)	1,031,887 (610,558–1,823,595)	968,775 (456,735–2,107,910)	1,116,610 (481,984–2,326,629)

[a] Models used: CU, strict clock with uniform priors; CP, strict clock with pure birth priors; LOG, uncorrelated relaxed clock lognormal model with pure birth priors; EXP, uncorrelated relaxed clock exponential model with pure birth priors.
[b] Divergence time between the genera *Sus* and *Phacochoerus*.
[c] Coalescent time of the different species.
[d] Time expressed in years. In parenthesis is the 95% high posterior density.

than previous estimates (Randi et al., 1996) and is more precise, since it takes into account errors both in topology and branch lengths and in molecular clock calibration. Interestingly, the TMRCA of the genus seems to be quite recent, approximately 1,269,712 years B.P. (734,328–2,340,764 95% high posterior density): this fits the high level of polyphyly encountered. We also noted that the coalescent time estimate of *S. scrofa* is similar to that inferred by Giuffra et al. (2000) and completely different from the much younger estimate of Kim et al. (2002).

3.2. Distance analyses

Distance analyses for species attribution were initially performed using the TN93 model of nucleotide substitution (Tamura and Nei, 1993), following the suggestions of Steinke et al. (2005). We first computed the average within-species distance and then used each sequence as a query (unknown) sequence, to determine to which species it could be assigned using different threshold levels. Interestingly, every query sequence could be attributed to any *Sus* species using a threshold only five times the average within-species distance. This means that cytochrome b does not perform well as an identification tool for *Sus*. Even more worrisome, when the *P. africanus* was used as query sequence, it was not recognized as a new species using the tenfold threshold proposed by Hebert et al. (2004). In other words, cytochrome b poses problems both of identification and of the discovery of new species. This is not surprising for three reasons: the abundant level of polyphyly found, the relatively young age of the genus, and the low level of intra-generic distances among cytochrome b *Sus* sequences compared with other mammalian genera (see Johns and Avise, 1998). However, it is worth noting that, when we performed the same analyses with the GTR+Γ nucleotide model of evolution (with substitution and rate heterogeneity parameters estimated by maximum likelihood on topology *a* of Fig. 2 using PAML; Yang, 1997), we found the same results in identification analysis, but it was possible to recognize *P. africanus* as a new species. This finding highlights the importance of model choice also in the barcoding field, especially for growing genetic divergences.

It should be noted that the distance approach used in species identification, although very popular in the growing field of barcoding (Hebert et al., 2004; Hajibabaei et al., 2006), is free from any population genetics assumption; in addition the threshold used is arbitrary and not dependent on the effective population size of the species examined. However, coalescent theory states that the expected number of pairwise differences between two sequences belonging to the same population follows a modified geometric distribution with parameter $1/(1+\theta)$, θ being twice the product of effective population size and mutation rate in haploid markers (Hein et al., 2004). Consequently, it is possible to derive a threshold (θ dependent) of the number of expected pairwise differences using its cumulative distribution function (Nielsen and Matz, 2006). We thus estimated θ in *S. celebensis*, and further assumed that all the other *Sus* species shared the same value (clearly an approximation). We chose *S. celebensis* for two reasons: the good sample size available and its restricted habitat (which makes the coalescent assumption of a random mating population reliable). We analyzed the 20 *S. celebensis* D-loop sequences published by Larson et al. (2005) to estimate θ (under the GTR+Γ model) and then scaled it to account for the different mutation rate of cytochrome b compared with the D-loop (scaling achieved through comparison of D6 D-loop sequences coalescence and D6 cytochrome b sequences coalescence; see Table S2 for samples used). The maximum *a posteriori* value of θ was used to determine the 95% probability interval that a query sequence could be attributed to a test species, given the observed pairwise difference. According to this approach, each query sequence could be attributed to any *Sus* species, confirming the identification problem found above; instead, warthog was correctly recognised as a new species.

In summary, cytochrome b cannot discriminate the various *Sus* species if a barcoding approach is used (although the value of this criticism is limited to cytochrome b). Specif-

ically, it is not possible to determine if New Guinea pig mitochondrial lineages are of *S. celebensis* or *S. scrofa* origin, therefore Y chromosome or autosomal sequences are needed.

Acknowledgment

This study received financial support from the Italian Ministry of University and Scientific Research (P.R.I.N. 2001–2003).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2007.05.025.

References

Alves, E., Ovilo, C., Rodriguez, M.C., Silio, L., 2003. Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and other domestic and wild pig populations. Anim. Genet. 34, 319–324.

Bollback, J.P., 2002. Bayesian model adequacy and choice in phylogenetics. Mol. Biol. Evol. 19, 1171–1180.

Drummond, A.J., Ho, S.Y., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. Plos. Biol. 4 (5), e88.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences (with discussion). Stat. Sci. 7, 457–511.

Gelman, A., Meng, X.-L., Stern, H., 1996. Posterior predictive assessment of model fitness via realized discrepancies. Stat. Sinica 6, 733–807.

Gerloff, U., Schlötterer, C., Rassmann, K., Rambold, I., Hohmann, G., Fruth, B., Tautz, D., 1995. Amplification of hypervariable simple sequence repeats (microsatellites) from excremental DNA of wild living Bonobos (*Pan paniscus*). Mol. Ecol. 4, 515–518.

Giuffra, E., Kijas, J.M., Amarger, V., Carlborg, O., Jeon, J.T., Andersson, L., 2000. The origin of the domestic pig: independent domestication and subsequent introgression. Genetics 154, 1785–1791.

Goldman, N., 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. 36 (2), 182–198.

Grillo, G., Attimonelli, M., Liuni, S., Pesole, G., 1996. CLEANUP: a fast computer program for removing redundancies from nucleotide sequence databases. Comput. Appl. Biosci. 12, 1–8.

Groves, C.P., 1981. Ancestors for the Pigs: Taxonomy and Phylogeny of the Genus *Sus*. Australian National University Press, Canberra.

Groves, C.P., 1997. Taxonomy of the wild pigs (*Sus*) of the Philippines. Zool. J. Linn. Soc. 120, 163–191.

Groves, C.P., Grubb, P., 1993. The Eurasian suids, *Sus* and *Babyrousa*—taxonomy and description. in: Oliver, W.L.R. (Ed.), Pigs, Peccaries, and Hippos. Status Survey and Conservation Action Plan. International Union for the Conservation of Nature, Gland, pp. 107–111.

Groves, C.P., Shaller, G.B., Amato, G., Khounboline, K., 1997. Rediscovery of the wild pig *Sus bucculentus*. Nature 386, 335.

Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W., Hebert, P.D., 2006. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. U.S.A. 103 (4), 968–971.

Hebert, P.D., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of birds through DNA barcodes. Plos. Biol. 2 (10), e312.

Hein, J., Schierup, M., Wiuf, C., 2004. Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory. Oxford University Press, Oxford.

Ho, S.Y., Jermiin, L., 2004. Tracing the decay of the historical signal in biological sequence data. Syst. Biol. 53 (4), 623–637.

Johns, G.C., Avise, J.C., 1998. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. Mol Biol. Evol. 15 (11), 1481–1490.

Kass, R.E., Raftery, A.E., 1995. Bayes factors. J. Am. Stat. Assoc. 90, 773–795.

Kim, K.I., Lee, J.H., Li, K., Zhang, Y.P., Lee, S.S., Gongora, J., Moran, C., 2002. Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism. Anim. Genet. 33, 19–25.

Larson, G., Dobney, K., Albarella, U., Fang, M., Matisoo-Smith, E., Robins, J., Lowden, S., Finlayson, H., Brand, T., Willerslev, E., Rowley-Conwy, P., Andersson, L., Cooper, A., 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science 307, 1618–1621.

Laud, P.W., Ibrahim, J., 1995. Predictive model selection. J. Roy. Stat. Soc. B 57, 247–262.

Lucchini, V., Meijaard, E., Diong, C.H., Groves, C.P., Randi, E., 2005. New phylogenetic perspectives among species of South-East Asian wild pigs (*Sus* sp.) based on mtDNA sequences and morphometric data. J. Zool., Lond. 266, 25–35.

Nielsen, R., Matz, M., 2006. Statistical approaches for DNA barcoding. Syst. Biol. 55 (1), 162–169.

Nylander, J.A., 2004. MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Nylander, J.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53 (1), 47–67.

Rambaut, A., Drummond, A.J., 2004. TRACER. University of Oxford, Oxford.

Randi, E., Lucchini, V., Diong, C.H., 1996. Evolutionary genetics of the Suiformes as reconstructed using mtDNA sequencing. J. Mamm. Evol. 3, 163–193.

Robins, J.H., Ross, H.A., Allen, M.S., Matisoo-Smith, E., 2006. Taxonomy: *Sus bucculentus* revisited. Nature 440, E7.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19 (12), 1572–1574.

Schmidt, H.A., Strimmer, K., Vingron, M., von Haeseler, A., 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18 (3), 502–504.

Steinke, D., Vences, M., Salzburger, W., Meyer, A., 2005. TaxI: a software tool for DNA barcoding using distance methods. Philos. T. Roy. Soc. B 360, 1975–1980.

Suchard, M.A., Weiss, R.E., Sinsheimer, J.S., 2005. Models for estimating Bayes factors with applications to phylogeny and tests of monophyly. Biometrics 61 (3), 665–673.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.

Welch, J.J., Fontanillas, E., Bromham, L., 2005. Molecular dates for the "Cambrian explosion": the influence of prior assumptions. Syst. Biol. 54 (4), 672–678.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z., Rannala, B., 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst. Biol. 54 (3), 455–470.