Dating of divergences within the Rattus genus phylogeny using whole mitochondrial genomes

Judith H. Robins a,*, Patricia A. McLenachan b, Matthew J. Phillips c, Lauren Craig d, Howard A. Ross e, Elizabeth Matisoo-Smith a

a Department of Anthropology and Allan Wilson Centre for Molecular Ecology and Evolution, The University of Auckland, Auckland, New Zealand
b Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand
c Centre for Macroevolution and Macroecology, School of Botany and Zoology, Australian National University, Canberra, Australia
d Dunedin School of Medicine, University of Otago, Dunedin, New Zealand
e Bioinformatics Institute and School of Biological Sciences, The University of Auckland, Auckland, New Zealand

Abstract

The timing and order of divergences within the genus Rattus have, to date, been quite speculative. In order to address these important issues we sequenced six new whole mitochondrial genomes from wild-caught specimens from four species, Rattus exulans, Rattus praetor, Rattus rattus and Rattus tanezumi. The only rat whole mitochondrial genomes available previously were all from Rattus norvegicus specimens. Our phylogenetic and dating analyses place the deepest divergence within Rattus at ~3.5 million years ago (Mya). This divergence separates the New Guinean endemic R. praetor lineage from the Asian lineages. Within the Asian/Island Southeast Asian clade R. norvegicus diverged earliest at ~2.9 Mya. R. exulans and the ancestor of the sister species R. rattus and R. tanezumi subsequently diverged at ~2.2 Mya, with R. rattus and R. tanezumi separating as recently as ~0.4 Mya. Our results give both a better resolved species divergence order and diversification dates within Rattus than previous studies.

1. Introduction

Whole mitochondrial genomes, as they have become more readily available, are being used to infer phylogenetic relationships among a range of organisms, e.g., mammals (Arnason et al., 2004a; Cao et al., 2000; Gibson et al., 2005; Phillips et al., 2006; Phillips and Penny, 2003), fish (Arnason et al., 2004b), birds (Harrison et al., 2004), and worms (Bleidorn et al., 2007). Until now the only whole mitochondrial genomes available for Rattus were from 13 Rattus norvegicus samples, and of these only three were from wild-caught individuals, with the remainder being from highly inbred laboratory strains (Schlick et al., 2006). Here we present six new whole mitochondrial genomes from four species of wild Rattus. We used these, together with sequences from the three wild R. norvegicus, a mouse (Mus musculus) and a vole (Volemys kikuchii), all obtained from GenBank, to date divergences within Rattus.

Rattus taxonomy is complex: Musser and Carleton (2005) described 65 species within the genus (subfamily Murinae, family Muridae, order Rodentia). Identification of individuals, even by experts, is often difficult (Taylor et al., 1982) and the evolutionary history of the genus is not fully resolved (Aplin, 2006). Most of the phylogenetic studies to date (e.g., Adkins et al., 2001, 2003; Jansa and Weksler, 2004; Michaux and Catzeflis, 2000; Michaux et al., 2001; Steppan et al., 2004, 2005) have investigated deeper divergences within Rodentia leaving the age of Rattus, and the timing of divergences within the genus, largely unknown.

Watts and Baverstock (1995), after extensive studies using the immunological distance method of microcomplement fixation of albumin, suggested that the evolutionary history of murine rodents was as follows. Murines arose about 20 Mya (million years ago) in southern Asia and began dispersing widely. By about 8 Mya extensive radiations had occurred within Southeast Asia, Australia, New Guinea, the Philippines and Africa. These were followed by rapid speciation but little further dispersal, except for the genus Rattus which arose in the Greater Sunda Islands much more recently and spread to New Guinea, Australia and continental Asia.

Using L1 (LINE-1, long interspersed repeated) retrotransposon elements to study the evolution of the Rattus lineage, Verneau et al. (1998) dated the divergence of the Rattus sensu stricto lineage from the other murids at 5.5–7.5 Mya. This was followed by two episodes of speciation, one at about 2.7 Mya and a second at about 1.2 Mya. Based on this work, together with that of Watts and Baverstock (1995), Aplin et al. (2003) noted that the date of about 2–3 Mya for the origin for Rattus was very recent given the size and diversity of the genus.

Jansa et al. (2006) investigated the interrelationships and the timing of diversification of the endemic rodents of the Philippines.
They used sequence data from the nuclear interphotoreceptor retinoid binding protein (IRBP) gene and the mitochondrial cytochrome b gene to construct a data set of 60 species which included four Rattus species. While their use of both mitochondrial and nuclear genes increased phylogenetic resolution, the authors used the nuclear gene alone for estimating divergence times as they observed that including the mitochondrial signal compromised the time estimates. They inferred from their results that the earliest murines were in the Philippines 15–18 Mya, in the Australo-Papuan region 8–10 Mya and that the origin of the clade containing the Rattus species was about 3–4 Mya.

Robins et al. (2007) used mitochondrial DNA (mtDNA) to investigate further the divergences among Rattus species from Island Southeast Asia, New Guinea, the Solomon Islands and islands further East in the Pacific although the primary purpose of their study was to identify Rattus species from these areas. They included some 19 species and built a phylogeny from three mtDNA regions the results of which generally supported the taxonomy of Musser and Carleton (2005). The molecular phylogeny resolved two major clades that divided the rats of the Australo-Papuan region from those of Asian and Island Southeast Asian origin. The depth and complexity of the clades was suggestive of divergences around 3 Mya within Rattus which, if borne out, would support an earlier entry of Rattus to Australia and New Guinea than the 1–2 Mya suggested by others (e.g., Aplin et al., 2003; Watts and Aslin, 1981).

In order to investigate more fully the timing of divergences within Rattus we sampled some of the groups identified in Robins et al. (2007) with particular emphasis on the commensal rats that became widespread in the Pacific. Rattus exulans occurs from Southeast Asia to the far reaches of the Pacific and is known to have been transported by prehistoric peoples. Matsuo-Smith and Robins (2004) identified four main lineages of R. exulans each with markedly different distributions centred, respectively, on the Philippines, Thailand, New Guinea and the islands of the wider Pacific. The complete mitochondrial genome of an example from each of three of these lineages was sequenced, as part of our ongoing studies of human migration in the Pacific. Our samples from the Philippines, unfortunately, yielded DNA of insufficient quality to allow the fourth lineage to be fully sequenced. Rattus rattus and Rattus tanezumi are closely related sister species. R. tanezumi is widespread in Asia and Island Southeast Asia (Musser and Carleton, 2005) while R. rattus dispersed to Europe by about the 3rd century AD and was carried worldwide via European ships from about 1700 AD onwards (Innes, 1990). R. norvegicus spread to Europe reaching western European cities early in the 18th century and, like R. rattus, it was subsequently transported around the world aboard European ships (Innes, 1990). In this study Rattus praetor represents the diverse Australo-Papuan lineage. This entire group is one of the least dispersed with Rattus praetor, an endemic of New Guinea, being unusual as it also occurs in the Solomon Islands, and has been found as sub-fossil remains recovered from archaeological sites in the Pacific as far east as Fiji. It was almost certainly transported beyond New Guinea by prehistoric peoples (White et al., 2000).

The results from the previous studies using LINE-1, IRPB and segments of mtDNA have left a number of issues unresolved. Our analysis of whole mitochondrial genomes, however, does reliably address these issues of dating and divergence within Rattus.

2. Methods

2.1. Samples

Whole mitochondrial genomes were sequenced from six rats comprising three R. exulans and one each of R. rattus, R. tanezumi and R. praetor. Since all our rats were wild-caught we chose to include in our analyses the GenBank sequences from the three wild-caught R. norvegicus (Schlick et al., 2006) rather than from any of the ten highly inbred strains from that study. Additionally sequences from GenBank from both mouse and vole were included as outgroups. Sample locations and GenBank Accession Nos. are detailed in Table 1.

In our initial analyses, we included the first complete rat mitochondrial genome sequenced (Gadaleta et al., 1989) which was from a Wistar strain R. norvegicus (NC_001665.1, derived from X14848). This was the GenBank reference sequence until July 2006. We encountered problems aligning this sequence because of many unique indels, especially in the 16S rRNA region. Subsequently we used it in only one dating analysis to see how it influenced the divergence estimates.

2.2. Sequence acquisition and selection

Total genomic DNA was extracted from liver or muscle tissue for all six rats using the High Pure PCR Template Preparation Kit (Roche). Primers were chosen for PCR from an extensive database of primers for mammals and birds developed at the Allan Wilson Centre for Molecular Ecology and Evolution, at Massey University, Palmerston North, New Zealand. In the absence of other Rattus mitochondrial genomes, that of R. norvegicus from GenBank Accession No. NC_001665.1, was used for initial searches to find suitably homologous primers from our primer data base. Additional primers were designed to complete the genomes when primers of sufficient similarity were not available. A full set of amplification primers for the rat genomes are in Table S1 (online supplementary information).

Two overlapping long-range PCR products, ranging from 9 to 11.5 kb in length, were amplified per specimen, to cover the entire mitochondrial genome. Primers used for the generation of long-range products and the sizes of the fragments are given in Table

Table 1

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Label in Fig. 1</th>
<th>GenBank accession number</th>
<th>Origin of specimen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volemys tikuchii</td>
<td>Vole</td>
<td>AF348082</td>
<td>Taiwan</td>
</tr>
<tr>
<td>Mus musculus</td>
<td>Mouse</td>
<td>NC_001569</td>
<td>Wirgote, Papua New Guinea</td>
</tr>
<tr>
<td>Rattus praetor</td>
<td>praetor PNG</td>
<td>EU273708</td>
<td>Great Barrier Is, New Zealand</td>
</tr>
<tr>
<td>Rattus exulans</td>
<td>exulans NZ</td>
<td>EU273711</td>
<td>Thailand</td>
</tr>
<tr>
<td>Rattus exulans</td>
<td>exulans Thailand</td>
<td>EU273710</td>
<td>Yuro, Papua New Guinea</td>
</tr>
<tr>
<td>Rattus rattus</td>
<td>rattus NZ</td>
<td>EU273709</td>
<td>Titirangi, New Zealand</td>
</tr>
<tr>
<td>Rattus tanezumi</td>
<td>tanezumi Japan</td>
<td>EU273712</td>
<td>Amami Is, Japan</td>
</tr>
<tr>
<td>Rattus norvegicus</td>
<td>norvegicus USA</td>
<td>DQ673916</td>
<td>Milwaukee, United States of America</td>
</tr>
<tr>
<td>Rattus norvegicus</td>
<td>norvegicus Denmark</td>
<td>AJ428514</td>
<td>Copenhagen, Denmark</td>
</tr>
<tr>
<td>Rattus norvegicus</td>
<td>norvegicus Japan</td>
<td>DQ673917</td>
<td>Tokyo, Japan</td>
</tr>
</tbody>
</table>

Origin of specimen refers to the location where the specimen was collected.
S2 (online supplementary information). The sequences of these primers are provided in Table S3 (online supplementary information). The protocols used were those for the Expand Long Template PCR System (Roche). The long-range products were then used as template for shorter PCR products (800 bp–2 kb). This procedure seeks to avoid amplifying nuclear copies of the mitochondrial genome (numts), which are smaller in length. Numts longer than 6 kb are rare in mammals although one of 7.9 kb has been characterised in the domestic cat (Lopez et al., 1994) and of eighty two numts longer than 1 kb within the human genome described by Hazzani-Covo et al. (2003) only one, at 8.8 kb was longer than 6 kb. Short-range PCR was carried out using a standard 3–step protocol: 94 °C, 3 min, 35 cycles of: 94 °C, 30 s, 50–60 °C, 30 s, 72 °C, 1 min followed by 1 cycle of 72 °C, 5 min and a hold at 15 °C. Each PCR product was purified for sequencing either through a sephacryl column (Microspin S300 from G E Healthcare) or was gel cut and cleaned through a column (High Pure PCR Product Puri fication Kit by Roche). The prepared products were sequenced on an ABI3730 capillary sequencer (Massey University, Albany, New Zealand). All the raw sequences were assembled, edited and aligned using the software package Sequencher14 (Gene Codes Corporation).

The dataset includes 12 mitochondrial genome proteins, rRNA and tRNA coding sequences, totalling 14,471 nucleotides after manual alignment in Se-Al (Rambaut, 1996) and the exclusion of nucleotide sites with ambiguous alignment. The D-loop was excluded from the dataset because of alignment difficulties due to its high variability. Also, typically for mitochondrial genome analyses, NDS was omitted because unlike the other 12 proteins it is encoded on the mitochondrial L-strand which has a different nucleotide composition from the H-strand. Assuming an homogeneous substitution process across whole mitochondrial genome sequences can lead to among-site rate heterogeneity being underestimated, even under a gamma distribution (Phillips and Penny, 2003). Hence, for Bayesian inference and maximum likelihood hypothesis testing the data were partitioned into three sets: protein codon positions 1 and 2, protein codon position 3 and RNA.

2.3. Phylogenetic methods

MODELEST 6.0 (Posada and Crandall, 1998) was used to identify optimal substitution models as the most general model identified by either the Akaike Information Criterion (AIC) or the hierarchical likelihood ratio test. For each partition this was found to be GTR + I + G, which was subsequently used to model each of the partitioned maximum likelihood (ML) and Bayesian inference analyses. Under MODELEST only stationary models of sequence substitution can be evaluated, however, one condition of stationarity is that base composition does not vary among taxa. To test this, the relative proportions of the four bases at all variable sites in the dataset were examined using the chi-square test in PAUP* 4.10b10 (Swofford, 2002). If the ingroup is essentially homogeneous for nucleotide composition, then alternative positions within Rattus will be affected approximately equally by any heterogeneity in the outgroup. As such, we focus on composition homogeneity in the mammalian ingroup because affinities among Rattus are our main concern and the outgroup status relative to these of mouse and vole is well established. The chi-square test revealed that compositional heterogeneity was a feature of the data (P < 0.001).

To identify the source of the compositional heterogeneity, we repeated the chi-square tests for the first and second protein positions, the third protein codon position, and the RNA sequences. Only in the case of the third protein codon position was there significant heterogeneity (P < 0.001). The respective P-values for protein codons 1 and 2 and the RNA data were 0.232 and 0.275. When the third protein codon position was recoded using RY-coding there was no significant heterogeneity (P = 0.17). Hence, we RY-coded the third protein codon positions in all subsequent phylogenetic analyses. An additional benefit of RY-coding the third positions is increased phylogenetic signal retention, which we assessed by computing the “stemminess” (Fiala and Sokal, 1985) of trees estimated from alignments with and without RY-coding. The stemminess is the uncorrected distance attributed to internal branch-lengths as a proportion of overall branch-length and at 0.340 is 14% higher for the RY-coded third positions than under standard nucleotide coding (0.299).

Maximum parsimony (MP) with bootstrapping (1000 replicates) was performed on the concatenated sequences using PAUP*. Bayesian inference tree building was conducted on partitioned data using MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001) and BEAST (Drummond et al., 2006). The partitioned data set was used in the MrBayes analyses with unlinked substitution models and branch-length rate multipliers among partitions. Three Markov chain Monte Carlo (MCMC) chains for each of two independent runs proceeded for 2,500,000 generations with trees being sampled every 1500 generations. The burnin length (150,000 generations) was determined by examination of −ln L versus MCMC generation plots and sampling efficiency (within Tracer 1.2 (Rambaut and Drummond, 2003)).

Support among the alternative tree topologies was examined using the Kishino–Hasegawa test (Kishino and Hasegawa, 1989) and approximately unbiased (AU) tests (Shimodaira, 2002) within the CONSEL program (Shimodaira and Hasegawa, 2001), with the sitewise log likelihoods combined from the GTR + I + G4 (CF87 + I + G4 for RY-coded data) maximum likelihood (ML) analyses run in PAUP* for each of the separately run partitions.

In BEAST, the three partitions were again separately modelled. Differences from the default settings included partitioning the equilibrium base frequencies and allowing ambiguities (i.e. RY-coded characters). Running BEAST for 5,000,000 generations with samples taken every 1000 ensured that all model parameters and times of the most recent common ancestors (tMRCA) were well sampled (ESS values > 100 in Tracer 1.2). The implementation of Bayesian inference in BEAST differs from that of MrBayes in a number of ways. MrBayes assumes an unrooted tree and no relationship between evolutionary rates along the different branches. In BEAST a “relaxed” phylogeny method is employed in which phylogeny and dates are simultaneously inferred. Rates along branches were distributed according to a lognormal distribution which provides greater flexibility than the exponential distribution option, in particular, the former model can better accommodate data that are nearly clock-like (Drummond et al., 2006).

We used the minimum and maximum age constraints of the mouse–rat divergence from Benton and Donoghue (2007) to calibrate the BEAST analysis. They stated that although the location of the split between Mus and Rattus is somewhat speculative, most current research suggests it is early in the evolution of Murinae but not basal in the divergence of the clade. We have used their recommendations of 12.3 Mya (the oldest record of Progonomys, the genus assumed to include the common ancestor of Mus and Rattus) and 11 Mya (based on records of the extinct genus Karnimata which is believed to be on the lineage leading to Rattus) as upper and lower demarcations, respectively, for the middle 95% of the normally distributed prior. In accord with all recent molecular and morphological interpretations the mouse plus rats clade was enforced relative to the vole. A uniform prior from 11 to 34 Ma for Muridae versus Cricetidae provides a calibration on the root that is very conservative at both the upper and lower bounds (see Steppan et al. (2004)).

The program r8s (Sanderson, 2003) was also used to estimate divergence times and the MrBayes tree provided an objective way to summarize branch-length differences into one tree for dating in r8s. The centre of the distribution for the mouse–rat diver-
gence date (11.65 Mya) was used in the r8s analyses in order to provide point estimates for the most recent common ancestors (tMRCAs). In r8s, the branch-lengths were smoothed using penalized likelihood. Although a likelihood ratio clock test rejected a uniform molecular clock ($P < 0.001$), we expect clock-testing to be very sensitive to violations when long sequences effectively reduce the influence of stochasticity. Indeed, the magnitude of apparent rate differences among branches is relatively small (see the phylogram provided in the supplementary information (Fig. S1)). Hence, for the cross-validation procedure within r8s, the smoothing value was almost a zero-gradient function of the smoothing parameter. We nominally chose a smoothing parameter of 100 and note that 100-fold lower or higher smoothing parameters made no more than 0.01 Ma difference to any of the node ages. We repeated the r8s analyses with the problematic R. norvegicus sequence (NC_001665.1) included.

3. Results

The six new rat mitochondrial genomes have been deposited in GenBank; see Table 1 for accession numbers. The overall lengths of the genomes were between 16293 and 16307 bp and all six have the standard mammalian gene order.

The whole mitochondrial genome sequence for the Wistar strain rat (NC_001665.1) was different from the other R. norvegicus sequences. Among the three wild R. norvegicus sequences the pairwise distance was ~0.5% whereas the distance between the Wistar rat and each of these was two to three times greater. The addition of the Wistar sequence to the alignment of the other three increased the alignment by 36 bases. In considering the 16S gene alone, the Wistar sequence, compared with those of the other rats, the vole and the mouse, introduced several regions of sequence with no obvious homology to that seen in other mammals, and several smaller insertions and deletions ranging from one to five bases. The most numerous of these were 28 one-base indels. These differences did not occur in any other sequences in our analysis including those of the outgroups. Since many of the indels occurred at sites that are otherwise highly conserved in mammals it is likely that these differences are due to sequencing error, a possibility with this sequence that was also noted by Schlick et al. (2006). Such error would be unsurprising given that the Wistar strain rat (NC_001665.1) was only the fourth mammalian mitochondrial genome to be completed (Gadaleta et al., 1989) and thus there was little to compare it with.

The MP tree and the Bayesian inference tree had the same topology (Fig. 1). The rats form a monophyletic group that has R. praetor as the deepest edge. The R. norvegicus group diverges at the next deepest node. This is followed by two more recent clades, the R. exulans group and the sister taxa, R. rattus and R. tanezumi. The deepest diverging members within R. norvegicus and R. exulans were, respectively, R. norvegicus USA and R. exulans NZ.

Our MrBayes analyses, which separately model the partitions while allowing them to share information on branch-lengths, likely provide the most accurate estimate of the tree. However, the associated Bayesian posterior probabilities (BPP) may confer over-confidence in tree support. This is because the assumption that stochastic error is accurately inferred from the posterior probability ratio of proposed versus current trees along the MCMC chain depends on the dubious assumption that the substitution models closely reflect the evolutionary processes under which the sequences have evolved (Felsenstein, 2004; Kelchner and Thomas, 2007). Shimodaira and Hasegawa (1999) showed that ML hypothesis tests such as the approximately unbiased (AU) and Kishino-Hasegawa (KH) tests can more closely reflect sampling error than do bootstrap values, which in turn are far more faithful than typically overconfident BPP values (Gontcharov et al., 2004; Suzuki et al., 2002).

Phylogenetic hypothesis testing in CONSEL compared the 15 trees (Table 2) that represent all alternative rooted topologies among the four major Rattus groups: R. praetor, R. norvegicus, R. exulans and R. tanezumi/R. rattus. Among these, a single tree topology (shown in Fig. 1) was significantly better than all others. All other topologies are rejected at $P \leq 0.036$ under the KH test and at $P \leq 0.043$ under the typically more conservative AU test. Thus, RH

![Figure 1](image)

**Fig. 1.** The Bayesian inference chronogram from BEAST. The grey bars represent the 95% credible intervals for tMRCAs from BEAST. The Bayesian posterior probability (BPP) and the MP bootstrap support for each node is given below. Node A represents the calibration. See Table 1 for a full description of samples.

<table>
<thead>
<tr>
<th>Trees</th>
<th>–ln L</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>KH</td>
</tr>
<tr>
<td>1</td>
<td>(Out, (Rpraet, (Rnov, (Rex, (Rrat, Rtan)))))) 35123.2&gt;</td>
<td>0.964 0.990</td>
</tr>
<tr>
<td>2</td>
<td>(Out, (Rnov, (Rpraet, (Rex, (Rrat, Rtan)))))) 15.4</td>
<td>0.036 0.043</td>
</tr>
<tr>
<td>3</td>
<td>(Out, (Rpraet, (Rnov, (Rex, (Rrat, Rtan)))))) 17.7</td>
<td>0.013 0.008</td>
</tr>
<tr>
<td>4</td>
<td>(Out, (Rpraet, (Rnov, (Rrat, Rtan)))))) 24.2</td>
<td>0.019 0.027</td>
</tr>
<tr>
<td>5</td>
<td>(Out, (Rpraet, (Rnov, (Rrat, Rtan)))))) 23.9</td>
<td>0.014 0.030</td>
</tr>
<tr>
<td>6</td>
<td>(Out, (Rpraet, (Rnov, (Rrat, Rtan)))))) 30.3</td>
<td>0.004 0.008</td>
</tr>
<tr>
<td>7</td>
<td>(Out, (Rrat, (Rnov, (Rpraet, Rnov)))))) 33.7</td>
<td>0.001 0.003</td>
</tr>
<tr>
<td>8</td>
<td>(Out, (Rnov, (Rex, (Rpraet, Rnov)))))) 39.4</td>
<td>0.005 0.007</td>
</tr>
<tr>
<td>9</td>
<td>(Out, (Rnov, (Rrat, (Rpraet, Rnov)))))) 40.3</td>
<td>0.004 0.019</td>
</tr>
<tr>
<td>10</td>
<td>(Out, (Rnov, (Rrat, (Rpraet, Rnov)))))) 40.7</td>
<td>0.004 0.009</td>
</tr>
<tr>
<td>11</td>
<td>(Out, (Rnov, (Rrat, (Rpraet, Rnov)))))) 41.0</td>
<td>0.003 0.010</td>
</tr>
<tr>
<td>12</td>
<td>(Out, (Rnov, (Rrat, (Rpraet, Rnov)))))) 43.1</td>
<td>0.001 0.004</td>
</tr>
<tr>
<td>13</td>
<td>(Out, (Rnov, (Rpraet, (Rrat, Rnov)))))) 43.9</td>
<td>0.001 0.001</td>
</tr>
<tr>
<td>14</td>
<td>(Out, (Rrat, (Rpraet, (Rnov, Rnov)))))) 44.3</td>
<td>0.001 0.001</td>
</tr>
<tr>
<td>15</td>
<td>(Out, (Rrat, (Rnov, (Rpraet, Rnov)))))) 47.8</td>
<td>&lt;0.001 0.001</td>
</tr>
</tbody>
</table>

Rnov and Rex each represent 3 taxa, and all possible arrangements of taxa within the R. norvegicus and R. exulans groups were considered when calculating –ln L scores. Only in the case of R. exulans was the arrangement of taxa important, with the relationship: (RexNZ, (RexThai, RexPNG)) being optimal. The –ln L scores given are for topologies with this arrangement within R. exulans, and with the arrangement (RnovUSA, (RnovJap, RnovDen)) within R. norvegicus. Out represents the outgroups mouse and vole. The –ln L differences from the best tree and their statistical significance under KH and AU tests are given. 

J.H. Robins et al. / Molecular Phylogenetics and Evolution 49 (2008) 460–466

463

Table 2 Log likelihood (–ln L) scores of the 15 tree topologies describing relationships among the four major Rattus groups: R. norvegicus (Rnov), R. exulans (Rex), R. praetor (Rpraet) and R. rattus & R. tanezumi (Rrat, Rtan).
divergences in our Fig. 1 is congruent with the order of divergences six new with RY-coded third protein codon positions and incorporating the sequence relative to the number of sequencing errors it includes. To disproportionately increase the branch-length of the erroneous sequences. Resultant model misspecification in analyses may tend for node H more than 3-fold from 0.12 to 0.43 Mya. It had a relation which increased from 2.78 to 2.95 Mya. This indicates that sequencing error can make a substantial difference to divergence esti-

4. Discussion

Our phylogeny (Fig. 1) based on whole mitochondrial genomes, with RY-coded third protein codon positions and incorporating the six new Rattus genomes, gives far better resolution of the relationships of the taxa included than any previous study. When the analysis was done without RY-coding, six alternative trees to the favoured tree could not be rejected in AU testing at $P < 0.05$ whereas with RY-coding all alternative trees could be rejected. The order of divergences in our Fig. 1 is congruent with the order of divergences inferred from analyses of LINE-1 (Verneau et al., 1998), the combined IRBP and cytochrome $b$ genes (Jansa et al., 2006) and the two mitochondrial genes plus D-loop (Robins et al., 2007). Our inferred phylogeny has strong support (Table 2), and shows an early divergence within Rattus ($\sim 3.5$ Mya) leading to the evolution of R. praetor and a lineage from which R. norvegicus, R. exulans, R. rattus and, R. tanezumi later diversified. Interestingly, despite their early divergence, R. praetor and its close relatives are quite restricted in their geographic distribution (Robins et al., 2007; White et al., 2000) whereas the other known commensal species are more widely distributed. This includes R. exulans, which was carried in the canoes of the first people to colonise Remote Oceania and was dispersed by them as far afield as Hawaii, Easter Island and the Chatham Islands off the east coast of New Zealand (Matiss-Smith and Robins, 2004).

Previous phylogenetic studies of Rattus species are consistent with the early divergence of the Australo-Papuan lineage. R. fuscipes, an Australian endemic, was found to diverge early in the LINE-1 study of Verneau et al. (1998) and both R. fuscipes and R. praetor occur within the Australo-Papuan clade of Robins et al. (2007). Some support for the early divergence of Rattus into the Australo-Papuan and Asian lineages was also suggested in the combined cytochrome $b$ and IRBP study of Jansa et al. (2006). Our estimates for this divergence time of $\sim 3.5$ Mya are consistent with the IRBP nuclear gene-based estimate of 3–4 Mya for the origin of the clade of Jansa et al. (2006) that contained all four of their Rattus species (R. exulans, R. everetti, R. praetor (sic), and R. tanezumi).

In the case of the LINE-1 study of Verneau et al. (1998), the deeper divergences among Rattus are younger than the equivalent ones in our phylogeny with the origin of the Australo-Papuan lineage at $\sim 2.7$ Mya compared with our $\sim 3.5$ Mya, the origin of the R. norvegicus lineage at $\sim 1.8$ Mya compared with our $\sim 2.9$ Mya, and R. exulans at $\sim 1$ Mya compared with our $\sim 2.2$ Mya. The LINE-1 results are based on very short sequences (215–315 bp) which makes them very susceptible to stochastic error but nevertheless the dates from Verneau et al. (1998) are still within our BEAST 95% credible limits and their divergence date for the lineages of R. rattus and R. tanezumi at $\sim 0.5$ Mya is very similar to our estimate.

Of their five calibration points, Jansa et al. (2006) used two within rodents: 12 Mya as the minimum date of origin for the murine crown group and 6 Mya for the appearance of Otomys. They used two methods to determine dates, the penalized likelihood approach in r8s (Sanderson, 2003) and a Bayesian method (Thorne and Kishino, 2002), and they analysed their data in three sets, cytochrome $b$ alone, IRBP alone and both genes combined. Their estimated divergence dates for the clade containing the Rattus samples were similar between the two dating methods but very different among the three data sets. Compared with using IRBP alone, analysis of the combined data set increased the age of the clade containing Rattus 3-fold, from 3–4 Mya to 10–13 Mya, and the use of cytochrome $b$ alone increased it 4-fold to 16–18 Mya. A similar pattern was seen at all nodes and the authors suggested that the cytochrome $b$ data severely violates the molecular clock among higher taxa. In contrast, our estimates with whole mitochondrial genomes resulted in dates that are congruent with those from the IRBP nuclear gene clearly indicating the advantages of estimating a different model of evolution for the third codon positions in mitochondrial sequences.

Our data suggest that the Australo-Papuan and Asian lineages separated at $\sim 3.5$ Mya. Given the restricted ranges of the modern Australo-Papuan members of Rattus, the founding lineages of Rattus are likely to have reached the Australia and New Guinea regions at a later date. Increased sampling within the Australo-Papuan clade will provide a minimum expected date for this migration. The phylogeny observed by Robins et al. (2007) suggests that the ancestral node of the Australo-Papuan clade, at approximately three quarters of the tree depth, is consistent with an age of $\sim 2.5$ Mya when compared with the divergence of the Rattus clades observed here at $\sim 3.5$ Mya. These dates suggest that the Australo-Papuan clade colonized the region sometime during a 1 Ma period between 2.5 and 3.5 Mya.

Our results also suggest that two further divergences in the Asian lineage rapidly followed the divergence of the Australo-Pap-

Table 3: Estimates in Mya for the most recent common ancestors: penalized likelihood point estimates from r8s and the median estimates plus 95% credible limits from BEAST

<table>
<thead>
<tr>
<th>Node (clade)</th>
<th>Penalized likelihood (in r8s)</th>
<th>BEAST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lower 95%</td>
<td>Median</td>
</tr>
<tr>
<td>A (calibration)</td>
<td>11.65</td>
<td>10.94</td>
</tr>
<tr>
<td>B</td>
<td>3.39</td>
<td>1.79</td>
</tr>
<tr>
<td>C</td>
<td>2.78</td>
<td>1.44</td>
</tr>
<tr>
<td>D</td>
<td>2.22</td>
<td>0.92</td>
</tr>
<tr>
<td>E</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>F</td>
<td>0.05</td>
<td>0.01</td>
</tr>
<tr>
<td>G</td>
<td>0.45</td>
<td>0.14</td>
</tr>
<tr>
<td>H</td>
<td>0.12</td>
<td>0.04</td>
</tr>
</tbody>
</table>
uan and Asian lineages, with first the split of *R. norvegicus* at ~2.9 Ma, followed by the split of *R. exulans* from *R. rattus* and *R. tanezumi* at ~2.2 Ma. These three divergences, nodes B–D in Fig. 1, followed in rapid succession, with little uncertainty as to their order, given the very strong support for the tree topology (Table 2). Overall, these ancient lineages of *Rattus* were established during a period of ~1.3 Ma.

The evolutionary events described above occurred in an area of the world subject to frequent changes in emergent land mass. The geologic history of the South East Asian and South West Pacific region is complex both tectonically and with respect to sea level change and has been extensively modelled by Hall (1998, 2001, 2002). He suggests that there were never continuous land links between Sundaland, comprising the Malay Peninsula, the Malay Archipelago islands of Sumatra, Java, Borneo and the surrounding smaller islands, plus the shallow seas of the Sunda shelf, and Australia but that areas of ephemeral land may have allowed island hopping. Hall (1998) postulated that beginning in the early Pliocene (~5 Mya), due to a combination of emergent small islands and changing sea levels, new dispersal pathways began developing in the region. An example he gave is a possible route connecting Thailand and New Guinea via the Sunda and Banda arcs, a pathway which is approximated today as through Malaysia, Indonesia, including the Lesser Sunda Islands and Ceram to the Bird’s Head peninsula of New Guinea. Although this route may have been available earlier than our median estimate of 3.5 Ma for the divergence of the Australo-Papuan and Asian lineages of *Rattus* it is within the 95% credible limits. Thus this path as it developed in stages and over time could have been a dispersal route for ancestral *Rattus* throughout all Southeast Asia and into both mainland Asia and the New Guinea region.

In addition to tectonic activity, sea level fluctuations, largely due to glaciation events, contributed to the development and the loss of possible migration routes. As discussed in Aplin (2006) there is evidence for sea ice surges from 4.7–4.3 Ma. This was followed by a warm period with associated higher sea levels until 3.95 Ma, then a long period of increased cooling with glacial events at 2.7 and 2.4 Ma. In the last 2 Ma there have been at least another 20 full cycles of glaciation and deglaciation. The climate change associated with these cycles of glaciation also brought large changes in vegetation with, for example in Northern Australia, rich rainforests in the interglacials giving way to dry sclerophyll forests (Aplin, 2006; van Oosterzee, 2006). The changing proportions of land and sea and the varying climatic conditions over the last 5 Ma have provided ample opportunity for dispersal, radiation, isolation and speciation in *Rattus*.

5. Conclusions

Our phylogenetic analysis of the six new *Rattus* whole mitochondrial genomes, together with the three already available from wild *R. norvegicus*, is concordant with the major division of *Rattus* identified by Robins et al. (2007) into those of the Australo-Papuan region and those of the Asian/Island Southeast Asian region. The deepest divergence at ~3.5 Ma is between *R. praeator*, representing the Australo-Papuan lineage, and the Asian lineage. Soon thereafter, the three main sub-lineages of the Asian clade diverge: *R. norvegicus* at ~2.9 Ma, followed by the split of *R. exulans* from *R. rattus* and *R. tanezumi* at ~2.2 Ma. Our results indicate that diversification within *Rattus* began earlier compared with the LINE-1 results (Verneau et al., 1998), but at a similar time compared with the IRPB results (Jansa et al., 2006). By partitioning the alignment into regions that are subject to very different substitution regimes and by RY-coding the third codon position in proteins, we obtain both good resolution of interspecies relationships among *Rattus* and well-supported dates that are much younger than those estimated from previous mitochondrial studies and are similar to nuclear estimates. These simple data treatments should be considered for future mitochondrial genome analysis.

Acknowledgments

We thank David Penny for advice and encouragement throughout the project and Peter Lockhart for advice on the manuscript. This research was funded by the Allan Wilson Centre for Molecular Ecology and Evolution. We are grateful to the South Australian Museum for providing us with tissues from four *Rattus* species.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.08.001.

References


