Genome Analysis

High mitogenomic evolutionary rates and time dependency

Sankar Subramanian[1,2], Dee R. Denver[3,4], Craig D. Millar[2], Tim Heupink[1], Angelique Aschrafi[3,5], Steven D. Emslie[6], Carlo Baroni[7] and David M. Lambert[1,3]

[1] Griffith School of Environment and the School of Biomolecular and Physical Sciences, Griffith University, 170 Kessels Road, Nathan, Qld 4111, Australia
[2] Allan Wilson Centre for Molecular Ecology and Evolution, School of Biological Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand
[3] Allan Wilson Centre for Molecular Ecology and Evolution, Institute of Molecular BioSciences, Massey University, Auckland, New Zealand
[4] Department of Zoology and Center for Genome Research and Biocomputing, Oregon State University, Corvallis, Oregon, USA
[5] Department of Cell Biology, The Scripps Research Institute, La Jolla, California 92037, USA
[6] Department of Biology and Marine Biology, University of North Carolina, Willmington, North Carolina, USA
[7] Dipartimento di Scienze della Terra, Universita' di Pisa, and Consiglio Nazionale Ricerche, Istituto di Geoscienze e Georisorse, Pisa, Italy

Using entire modern and ancient mitochondrial genomes of Adélie penguins (*Pygoscelis adeliae*) that are up to 44000 years old, we show that the rates of evolution of the mitochondrial genome are two to six times greater than those estimated from phylogenetic comparisons. Although the rate of evolution at constrained sites, including nonsynonymous positions and RNAs, varies more than twofold with time (between shallow and deep nodes), the rate of evolution at synonymous sites remains the same. The time-independent neutral evolutionary rates reported here would be useful for the study of recent evolutionary events.

Rates of molecular evolution

Rates of molecular evolution underpin much of modern evolutionary biology because they can be used, for example, to estimate the time of divergence between species and/or lineages and the age of morphological novelties. Molecular rates have generally been estimated using phylogenetic comparisons among multiple species, where levels of divergence have been calibrated using points in the fossil record [1–4] or major biogeographic events [5]. Recent studies have shown a much higher rate of evolution within species than that obtained using inter-species comparisons [6,7]. On the basis of on these observations, it has been suggested that rates of molecular evolution are time-dependent [8,9]. This apparent time dependency of evolutionary rates is potentially a result of two factors. First, the estimation of inter-species divergence at neutral sites such as synonymous positions and the hyper-variable region (HVR) is likely to suffer from the saturation effects of multiple substitutions. This would result in an underestimation of molecular rates over long periods [10–12]. Second, the presence of short-lived slightly deleterious mutations among populations of a species would lead to an overestimation of the actual substitutions over shorter timescales [12–14]. Although rate estimations based on within-species comparisons are not influenced by saturation effects, the presence of short-lived deleterious mutations still poses an obstacle to obtaining unbiased estimates of evolutionary rates for single species. Furthermore, phylogenetic methods cannot be used to infer intra-species rates owing to the lack of calibration points within a short timescale.

We have used serially preserved ancient DNA samples to resolve this impasse by sequencing the mitochondrial genomes of 12 modern and eight ancient Adélie penguins from Antarctica, including two samples dating back ~44 000 years [7,15]. These sequences enabled us to estimate precisely the overall mitochondrial genomic evolutionary rate and specific rates for ribosomal RNAs, transfer RNAs and synonymous and nonsynonymous positions of the protein-coding regions. These specific rates enabled us to examine the effects of time dependency on rates of molecular evolution. We have used three different analytical methods to estimate these molecular rates in mitochondrial genomes of Adélie penguins.

Estimating genomic rates of evolution using ancient mitochondrial DNA

We amplified and sequenced mitochondrial genomes using blood samples of modern Adélie penguins and from the sub-fossil bones of these birds from Antarctica (Figure 1). The size of the genome recovered from each of these ancient samples was 12.5–16.1 kb (Appendix BFigure S1 in the supplementary material online). The ages of these sub-fossil bones ranged from 250 to ~44 000 years, which were estimated using radiocarbon methods. The phylogenetic relationship of modern and ancient populations of Adélie penguins is shown in Appendix Bsupplementary Figure S2. First, to estimate the neutral evolutionary rate we analyzed only the synonymous positions (fourfold degenerate sites) of the mitochondrial protein-coding regions, which are known to be largely free from selection [16]. We used a Bayesian statistics-based Markov chain Monte Carlo (MCMC) method [17] and maximum likelihood [18] approaches to

Corresponding author: Lambert, D.M. (d.lambert@griffith.edu.au).

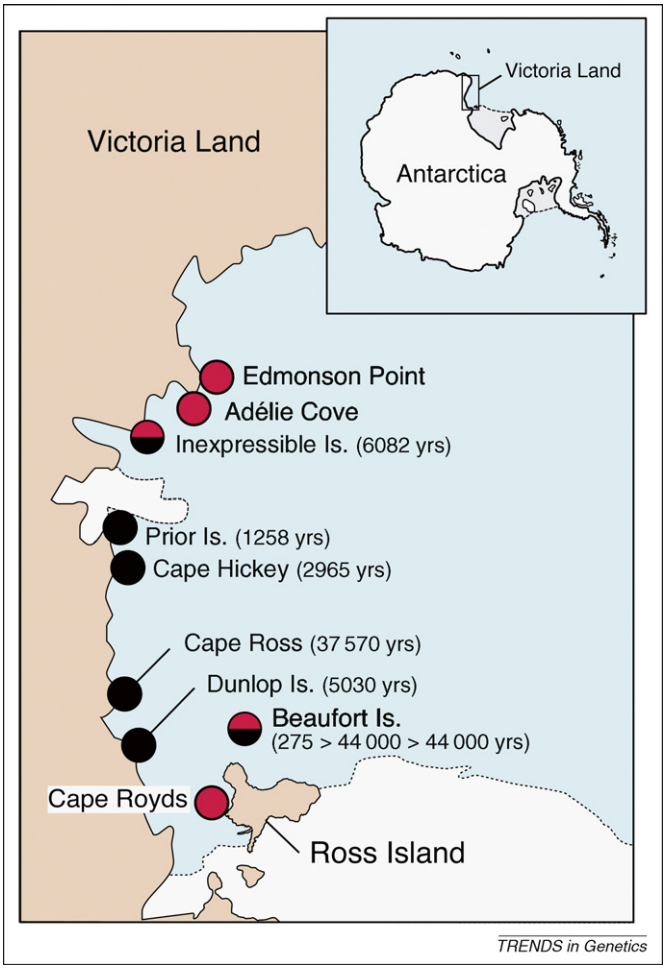

Figure 1. The location of ancient and modern samples of Adélie penguin used in this study. Ross Sea Coast colonies and Ross and Beaufort Island colonies were sampled. The red color indicates modern and black indicates the ancient samples.

estimate evolutionary rates. The Bayesian MCMC analysis showed that the mitogenomic rate of evolution at synonymous sites (μ_S) is 0.073 [95% highest posterior density (HPD) intervals 0.025–0.123 substitutions per site per million years (s/s/My)] (Figure 2). This rate estimate was obtained using a general time-reversible model of evolution. To account for rate variation among nucleotide positions, we used a discrete gamma distribution of six rate categories. The rate was then estimated assuming a constant population size and a relaxed uncorrelated log-normal molecular clock. The estimates remained almost the same when modifying a range of prior assumptions. To avoid any methodological bias, we also estimated the rate of evolution using a maximum likelihood method, which resulted in a rate of 0.062 s/s/My (CI 0.03–0.09). This rate is slightly less than, but comparable to, the rate estimated using the Bayesian MCMC approach (Figure 2). Furthermore, a similar rate was obtained using a simple distance-based serial UPGMA method [19] (0.072 s/s/My, CI 0.0–0.317).

Evolutionary rates from modern penguin samples

Next, we estimated the rate of mitogenome evolution in Adélie penguins using only the modern samples. We calculated the ratio of the mean pairwise divergences for noncoding sites at the HVR-I (d_H) and synonymous sites (d_S) using mitochondrial genomes recovered from 12 modern Adélie penguins. The rate for synonymous positions (μ_S) was estimated as $\mu_S = \mu_H . (d_S/d_H)$, where μ_H is the neutral mutation rate at the HVR-I region, which was previously estimated as 0.55 s/s/My [20]. Our current analysis shows that the synonymous sites from modern Adélie penguin mitochondrial genomes evolve at a rate (μ_S) of 0.054 (CI 0.031–0.090) s/s/My (Figure 2). This rate is not significantly different from those obtained using ancient DNA methods as the confidence intervals are almost the same.

Inferring tempo of evolution at constrained sites in the mitochondrial genome

The estimation of evolutionary rates at constrained sites (nonsynonymous sites and RNAs) is prone to large stochastic errors owing to the low divergence in population data. However, the average rate of evolution at constrained sites can be estimated using the synonymous evolutionary rate. The neutral theory defines the rate of evolution (K) at constrained sites as $K = \mu . f_0$ where μ is the mutation rate and f_0 is the fraction of neutral positions (or neutral mutations) [21]. Because the rate of evolution at synonymous sites (μ_S) of mitochondrial genes is known to be similar to the mutation rate (μ) and the divergence at these sites (d_S) largely reflects the accumulation of neutral mutations over time [16], the evolutionary rates in the constrained sites can be estimated as $K_X = \mu_S . (d_X/d_S)$, where K_X and d_X are the rate of evolution and divergence in constrained sites, respectively. The ratio (d_X/d_S) is actually the fraction of neutrally evolving positions (f_0). Using this approach, we estimated that the nonsynonymous rate of change for the mitochondrial genome of Adélie penguins is 0.005–0.007 s/s/My.

The synonymous molecular rate (0.054–0.073 s/s/My) differs from the nonsynonymous rate by more than 10 times (Figure 2). The mitochondrial tRNAs (0.015–0.020 s/s/My) evolve almost twice as fast as rRNAs (0.008–0.010 s/s/My). The rate of evolution in the complete mitochondrial molecule (excluding the control region) is 0.018–0.024 s/s/My, which is 1.8–2.4 times faster than the classical phylogenetic rate of 0.01 s/s/My that has been used for several decades [1,22]. A previous study on the mitochondrial genomes of the extinct mammoth also suggests that the rate based on internal calibrations (within mammoths) is ~1.6 times higher than that obtained using the external (i.e. mammoth–elephant) calibration [23].

By contrast, the synonymous substitution rate (0.054–0.073 s/s/My) estimated here is five to seven times higher than previous phylogenetic rate estimates [1–4] and significantly higher than those based on intra-specific comparisons within human (0.048–0.052 s/s/My) [14] and Neanderthal (0.036–0.042 s/s/My) [24] populations. These results clearly argue against the use of the classical 1% rate per lineage (or the '2% rule' as it is commonly known) to study the evolution or genetics of individual species.

Time dependency of molecular rates

It has previously been suggested that rates of molecular evolution are time-dependent and that rates estimated

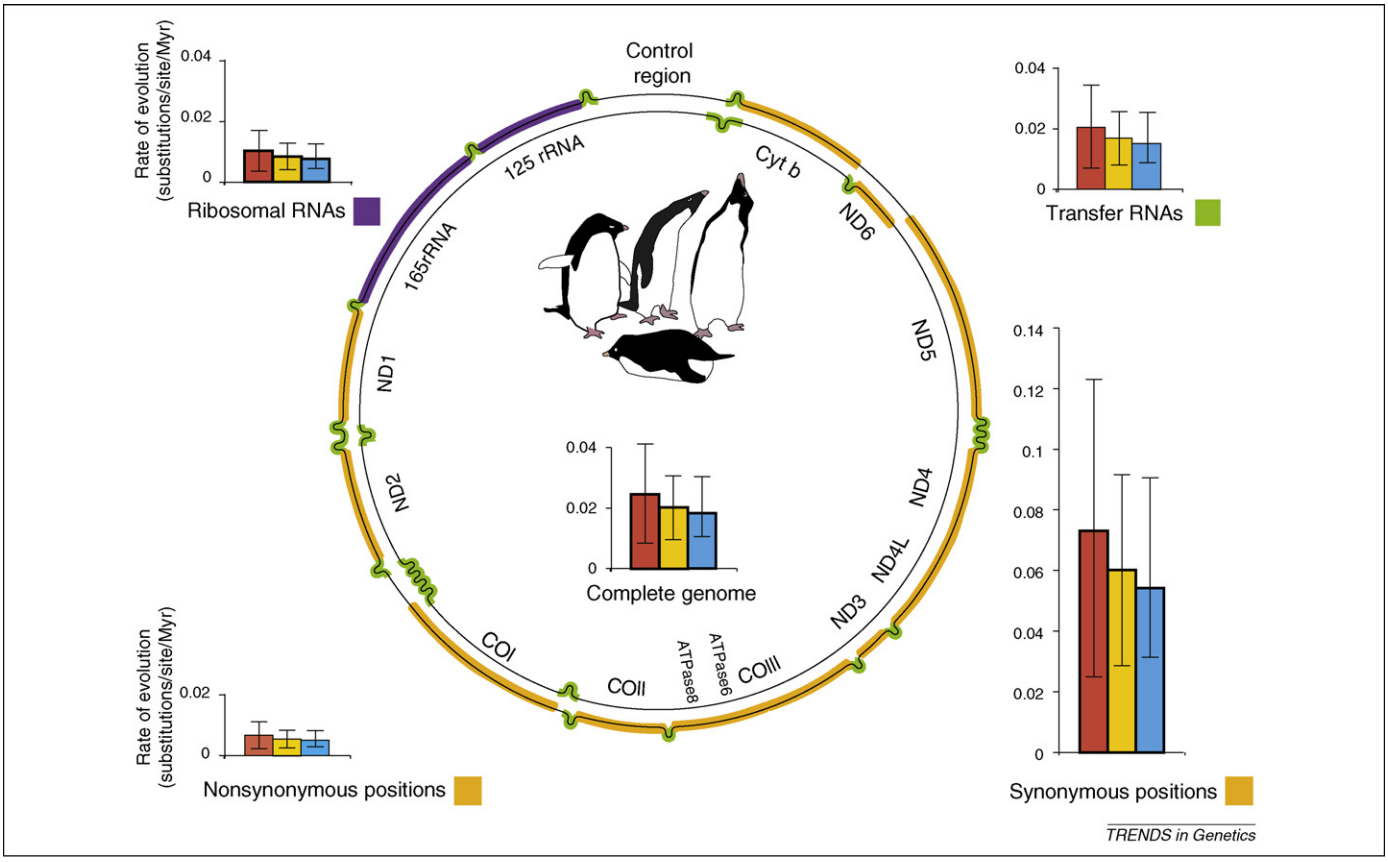

Figure 2. Mitogenomic rates of evolution in Adélie penguins. Using synonymous positions (fourfold degenerate sites) of the mitochondrial genomes of eight ancient and 12 modern Adélie penguins, the evolutionary rate at synonymous sites (μ_S) was estimated by Bayesian statistic-based MCMC (red column) and a maximum likelihood approach (yellow column). In addition, using only the modern genomes the synonymous rate was also inferred using the ratio of the average pairwise divergence in the synonymous positions (d_S) to that in the HVR (d_H) and using the mutation rate (μ_H) of the latter, as $\mu_S = \mu_H (d_S/d_H)$ where $\mu_H = 0.55$ (0.32–0.92) substitutions/site/million years as reported from a previous study [20] (blue column). The evolutionary rates in different mitochondrial regions (K_X) are estimated as $K_X = \mu_S (d_X/d_S)$ where d_X is the average pairwise distance obtained for the given region X (e.g. tRNA). The error bars show 95% HPD intervals.

over short timescales are higher than long-term evolutionary rates [6,9]. To detect any time dependency, we first determined the coalescence times of shallow and deep branches of modern Adélie penguin populations. Previous studies have shown the presence of two distinct lineages of Adélie penguins (i.e. Antarctic (A) and Ross Sea (RS) lineages) [7,15]. Using the complete mitogenomic evolutionary rate, we estimated the coalescence time between the lineages (A/RS) to be 62 000 years (62 Ky) and the coalescence time within individual lineages 18 Ky and 19 Ky for A and RS, respectively. We then estimated the root heights of the most recent common ancestor (μt_{MRCA}, substitutions per site) of A, RS and A/RS lineages using synonymous (third codon) positions and constrained (nonsynonymous sites + tRNAs + rRNAs) positions. The μt_{MRCA}s were estimated using the program MCMCcoal [25], and then using coalescence times the rate of evolution at constrained and synonymous sites for inter- and intra-lineages of Adélie penguins were computed.

Our examination of the UPGMA trees (Appendix Bsupplementary Figure S3) revealed that the node heights of A and RS lineages are much longer (with respect to the inter-lineage–A/RS height) for constrained sites (Appendix BFigure S3b) than those constructed using synonymous sites (Appendix BFigure S3a). The rate of evolution at constrained sites is more than twofold higher within A (0.02 s/s/My, HPD 0.014–0.033 s/s/My) and RS (0.017 s/s/My, HPD 0.011–0.034 s/s/My) lineages compared with that obtained for between A and RS lineages (0.008 s/s/My, HPD 0.007–0.01 s/s/My) (Figure 3). However, the evolutionary rates at synonymous sites estimated for intra- and inter-lineages are similar (Figure 3). In a recent study, a comparable pattern was observed in human populations [14]. The rate of evolution at third codon positions was found to be similar to that recorded within human populations (0.052 s/s/My) and between humans and chimpanzees (0.048 s/s/My). By contrast, the nonsynonymous rate of evolution within human populations (0.013 s/s/My) was 3.4 times higher than that estimated between humans and chimpanzees (0.0038 s/s/My). The rate difference (3.4 times) recorded for the hominid data is higher than that reported in this study (two times, Figure 3) because the hominid comparison is on a much deeper timescale (between species) than our comparison involving two lineages of the same species. Because multiple substitutions are not an issue for comparisons within a species or for closely related species (particularly for synonymous sites), there is no reason to expect a significant variation in the rate of neutral evolution at different timescales. These results suggest that the time dependency of evolutionary rates holds true only for the sites under selection and not for the neutral synonymous sites.

Furthermore, we examined the effect of time dependency on the rate for the mitochondrial HVR, which is also

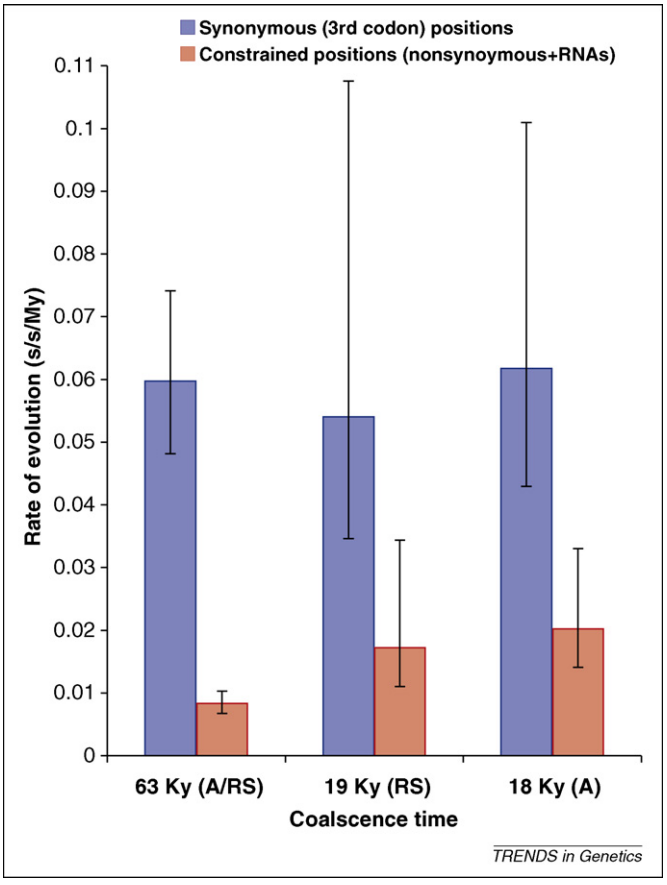

Figure 3. The rate of evolution for within and between the two lineages of modern Adélie penguins. The evolutionary rates were estimated using synonymous (third codon) positions and constrained positions (nonsynonymous sites + tRNAs + rRNAs). The coalescence times for within and between lineages were estimated by the program BEAST [17] using the rate obtained for the complete mitochondrial genome. The evolutionary rates for A and RS lineages and those between them (A/RS) were obtained using the respective coalescence times and root heights of the most recent common ancestor (μt_{MRCA}), which was estimated by MCMCcoal [25] using synonymous and constrained sites. The error bars are based on the 95% HPD intervals obtained from coalescence time estimation.

known to evolve under neutrality. The rate of evolution between lineages (A/RS) (0.496 s/s/My, HPD 0.399–0.615 s/s/My) was not significantly different from that within the A lineage (0.413 s/s/My, HPD 0.288–0.675 s/s/My), but the rate within the RS lineage (0.798 s/s/My, HPD 0.511–1.59 s/s/My) was higher than the inter-lineage rate. Since the rate of evolution within the RS lineage is based on only three sequences (Appendix Bsupplementary Figure S3) as opposed to nine sequences for the A lineage, the higher rate for the RS lineage could be simply because of stochastic errors. Furthermore, the well-known mutational hot spots in the HVR might also influence the rate estimation. For instance, Howell et al. [26] have recently shown that, while the rate of evolution of a large proportion of the HVR is comparable to that for synonymous sites, other sites mutate an order of magnitude faster than the remainder. Therefore, the rate difference observed within A and RS lineages is most likely to be a result of mutational processes and not from purifying selection or the effects of time dependency.

Concluding remarks

For decades, phylogenetic estimates of rates of molecular change have been widely used in evolutionary studies, but the field has recently been characterized by debates about the use of intra-specific rate estimates [6,7]. Previous studies of mitochondrial HVRs suggested much higher intra-specific evolutionary rates compared with classical phylogenetic rates estimated using multiple species. However, HVR rates cannot be used to infer molecular rates for mitochondrial genomes because this region evolves much faster than rest of the genome [12,27]. By contrast, the evolutionary rate for synonymous positions reported here is useful to infer the rates of constrained protein coding sites, tRNAs and rRNAs, which will help us understand the patterns of purifying selection acting at the population level. We have shown that the high neutral rate reported in this study is similar for shallow and deep branches. Therefore, this time-independent rate can be used to study the recent evolutionary histories of many avian species.

Acknowledgements

We thank Griffith University, the Australia–India Strategic Research Fund, the Allan Wilson Centre for Molecular Ecology and Evolution and the New Zealand Marsden Fund for financial support. The fieldwork was only possible with the assistance of the Antarctica New Zealand (DML and CDM), the Italian Antarctic Programme (CB) and the National Science Foundation and National Geographic Society (SE). The radiocarbon ages were provided by the Geochron Laboratory-Krueger Enterprise Inc., Cambridge, Massachusetts (conventional and AMS, GX-), the IsoTrace Radiocarbon Laboratory, Toronto (AMS, TO-), NOSAMS, Woods Hole Oceanographic Institute (AMS, OS-) and the New Zealand Institute of Geological and Nuclear Sciences, Lower Hutt, New Zealand (AMS, NZA-). We are grateful to reviewers for their constructive and helpful comments.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2009.09.005.

References

1 Shields, G.F. and Wilson, A.C. (1987) Calibration of mitochondrial DNA evolution in geese. *J. Mol. Evol.* 24, 212–217
2 Lovette, I.J. (2004) Mitochondrial dating and mixed support for the '2% rule' in birds. *Auk* 121, 1–6
3 Pereira, S.L. and Baker, A.J. (2006) A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. *Mol. Biol. Evol.* 23, 1731–1740
4 Packert, M. *et al.* (2007) Calibration of a molecular clock in tits (Paridae) – do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule? *Mol. Phylogenet. Evol.* 44, 1–14
5 Marko, P.B. (2002) Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. *Mol. Biol. Evol.* 19, 2005–2021
6 Ho, S.Y. and Larson, G. (2006) Molecular clocks: when times are a-changin'. *Trends Genet.* 22, 79–83
7 Lambert, D.M. *et al.* (2002) Rates of evolution in ancient DNA from Adélie penguins. *Science* 295, 2270–2273
8 Ho, S.Y.W. *et al.* (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol. Biol. Evol.* 22, 1561–1568
9 Penny, D. (2005) Evolutionary biology – relativity for molecular clocks. *Nature* 436, 183–184
10 Heyer, E. *et al.* (2001) Phylogenetic and familial estimates of mitochondrial substitution rates: study of control region mutations in deep-rooting pedigrees. *Am. J. Hum. Genet.* 69, 1113–1126
11 Nabholz, B. *et al.* (2008) Strong variations of mitochondrial mutation rate across mammals – the longevity hypothesis. *Mol. Biol. Evol.* 25, 120–130
12 Soares, P. *et al.* (2009) Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. *Am. J. Hum. Genet.* 84, 740–759

13 Kivisild, T. *et al.* (2006) The role of selection in the evolution of human mitochondrial genomes. *Genetics* 172, 373–387
14 Endicott, P. and Ho, S.Y. (2008) A Bayesian evaluation of human mitochondrial substitution rates. *Am. J. Hum. Genet.* 82, 895–902
15 Ritchie, P.A. *et al.* (2004) Ancient DNA enables timing of the Pleistocene origin and Holocene expansion of two Adélie penguin lineages in Antarctica. *Mol. Biol. Evol.* 21, 240–248
16 Yang, Z. and Nielsen, R. (2008) Mutation–selection models of codon substitution and their use to estimate selective strengths on codon usage. *Mol. Biol. Evol.* 25, 568–579
17 Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7, 214
18 Rambaut, A. (2000) Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. *Bioinformatics* 16, 395–399
19 Drummond, A.J. *et al.* (2002) Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* 161, 1307–1320
20 Millar, C.D. *et al.* (2008) Mutation and evolutionary rates in Adélie penguins from the Antarctic. *PLoS Genet.* 4, e1000209
21 Kimura, M. (1983) *The Neutral Theory of Molecular Evolution*, Cambridge University Press
22 Garcia-Moreno, J. (2004) Is there a universal mtDNA clock for birds? *J. Avian Biol.* 35, 465–468
23 Gilbert, M.T.P. *et al.* (2008) Intra-specific phylogenetic analysis of Siberian woolly mammoths using complete mitochondrial genomes. *Proc. Natl. Acad. Sci. U.S.A.* 105, 8327–8332
24 Briggs, A.W. *et al.* (2009) Targeted Retrieval and analysis of five neanderthal mtDNA genomes. *Science* 325, 318–321
25 Rannala, B. and Yang, Z.H. (2003) Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* 164, 1645–1656
26 Howell, N. *et al.* (2007) Relative rates of evolution in the coding and control regions of African mtDNAs. *Mol. Biol. Evol.* 24, 2213–2221
27 Quinn, T.W. (1992) The genetic legacy of Mother Goose–phylogeographic patterns of lesser snow goose *Chen caerulescens caerulescens* maternal lineages. *Mol. Ecol.* 1, 105–117

0168-9525/$ – see front matter © 2009 Published by Elsevier Ltd.
doi:10.1016/j.tig.2009.09.005 Available online 14 October 2009

Letters

Detecting new neurodegenerative disease genes: does phenotype accuracy limit the horizon?

David C. Samuels[1], David J. Burn[2,3] and Patrick F. Chinnery[3]

[1] Center for Human Genetics Research, Department of Molecular Physiology and Biophysics, Vanderbilt University Medical Center, Nashville, TN 37232, USA
[2] Clinical Ageing Research Unit, Institute for Ageing and Health, Campus for Ageing and Vitality, Newcastle-upon-Tyne, NE4 5PL, UK
[3] Mitochondrial Research Group, Institute for Ageing and Health & Institute of Human Genetics, The Medical School, Newcastle University, Newcastle-upon-Tyne, NE2 4HH, UK

Neurodegenerative diseases present a major health challenge to the ageing population, and most are thought to arise through a complex interplay between inherited genetic variation and environmental triggers. Although rare monogenic forms of common neurological disorders exist, these account for <5% of the total number of cases. Large-scale genome-wide association studies (GWAS) are starting to have some success in identifying the major risk alleles involved in several common neurodegenerative disorders [1]. However, as research moves to the next phase of GWAS, one needs to ask whether more emphasis should be placed on phenotypic accuracy, rather than simply increasing sample sizes.

The diagnosis of a late-onset neurological disease generally relies heavily on the clinical description provided by an experienced neurologist. Specific diagnostic tests are rare, and several autopsy case series have demonstrated a diagnostic error rate approaching ~10%, even in expert hands [2,3]. Diagnostic revision also occurs in ~1/3 of cases [4,5]. Phenotypic misclassification reduces the power to detect a statistical association between a phenotype and specific allele for a given sample size [6–9]. Although this is primarily a clinical issue, the genetics community should be concerned; it is also relevant for any human disease where the clinical classification is not 100% accurate [8,9].

There have been several approaches to try and deal with the issue of diagnostic inaccuracy in neurodegenerative diseases (Box 1). *In silico* modelling has shown that increasing the sample size counterbalances diagnostic error [10], but that the relationship between statistical power and diagnostic accuracy is not linear; in addition, the sample size required to generate reasonable power increases dramatically with reduced diagnostic accuracy [10]. For strong genetic effects, the precise diagnosis might not be a key issue. For example, even when 15% of cases are incorrectly classified as Alzheimer disease, a study of 500 cases and 500 controls would have >70% power to detect the well-established association with the ε4 *APOE* allele (Online Supplementary Material Figure S1). However, the detection of hitherto unknown modest disease associations at the whole-genome level presents a greater challenge [11]. For common genetic variants exerting a modest effect [where the genome relative risk (GRR) is 1.3], a diagnostic error rate of ~2% has little effect on statistical power (Figure 1a). However, >2% diagnostic error has a dramatic effect on power, especially when attention is drawn to lower-penetrance alleles (i.e. GRR ≤1.1), as proposed for many complex traits. This is further compounded when less frequent but equally plausible genetic variants (with a minor allele frequency ≤10%) are considered, which are highly sensitive to diagnostic errors (Figure 1b). Studies of rarer disease phenotypes (affecting <1 in 1000 adults)

Corresponding author: Chinnery, P.F. (p.f.chinnery@ncl.ac.uk).