



OPEN

Distance-dependent patterns of molecular divergences in tuatara mitogenomes

SUBJECT AREAS:
MOLECULAR EVOLUTION
GENETIC VARIATIONSankar Subramanian^{1*}, Elmira Mohandesan^{2*†}, Craig D. Millar³ & David M. Lambert¹Received
10 September 2014Accepted
2 February 2015Published
3 March 2015

Correspondence and requests for materials should be addressed to S.S. (s.subramanian@griffith.edu.au) or D.M.L. (d.lambert@griffith.edu.au)

* These authors contributed equally to this work.

† Current address: Institute of Population Genetics, University of Veterinary Medicine Vienna, Veterinärplatz 1, 1210 Vienna, Austria.

¹Environmental Futures Research Institute, Griffith University, Nathan 4111, Australia, ²Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, New Zealand, ³Allan Wilson Centre for Molecular Ecology and Evolution, School of Biological Sciences, University of Auckland, Private 92019, Auckland, New Zealand.

Population genetic models predict that populations that are geographically close to each other are expected to be genetically more similar to each other compared to those that are widely separate. However the patterns of relationships between geographic distance and molecular divergences at neutral and constrained regions of the genome are unclear. We attempted to clarify this relationship by sequencing complete mitochondrial genomes of the relic species Tuatara (*Sphenodon punctatus*) from ten offshore islands of New Zealand. We observed a positive relationship that showed a proportional increase in the neutral diversity at synonymous sites (dS), with increasing geographical distance. In contrast we showed that diversity at evolutionarily constrained sites (dC) was elevated in the case of comparisons involving closely located populations. Conversely diversity was reduced in the case of comparisons between distantly located populations. These patterns were confirmed by a significant negative relationship between the ratio of dC/dS and geographic distance. The observed high dC/dS could be explained by the abundance of deleterious mutations in comparisons involving closely located populations, due to the recent population divergence times. Since distantly related populations were separated over long periods of time, deleterious mutations might have been removed by purifying selection.

Tuatara are significant in terms of global biodiversity and the evolutionary history of reptiles¹. They are the sister taxon to the Squamata (a group that includes lizards, and snakes)² and are regarded as the most distinctive surviving reptilian genus. Although they resemble many lizard species, tuatara are not lizards at all. Instead, they are actually the last remaining member of a distinct order Sphenodontia, which has diverged from Squamata (all other reptiles) about 220 million years^{3,4}. This unique reptile was once widespread on the mainland of New Zealand before the arrival of humans. It was thought to have been driven to extinction by exotic mammals introduced around 800 years ago^{5,6}. Tuatara are now exclusively found on only 32 offshore islands around New Zealand.

It is well known that genetic similarity decreases with increasing distances between the habitats of individuals or population⁷. This is due to the balance between genetic drift and migration (gene flow). While migration typically introduces genetic variability, drift almost inevitably reduces it. However in the case of populations of a land vertebrate living on islands, the above relationship will largely be determined by genetic drift as the effects of migration is minimum. In the present study we examined this relationship using mitochondrial genomes of tuatara living in ten distinct islands that are 1–631 km separated. Although previous studies examined the relationship between geographic and genetic distances, their data were largely confined to neutral markers^{8–11}. It is unclear how purifying selection modulates the genetic diversities at constrained regions of genomes. To examine this we estimated diversities at neutral and constrained sites of tuatara mitochondrial genomes.

Results

Although tuatara live in 32 islands their population in most of the islands are very small ranging from tens to few hundreds with an exception of Stephens Island, which harbours >30,000 individuals (<http://www.doc.govt.nz/documents/science-and-technical/TSRP47.pdf>). As a result of fragmented and small population sizes tuatara are a highly endangered reptile and therefore obtaining samples is difficult. We managed to obtain tuatara samples from ten different islands. We sequenced the complete mitochondrial genomes of a single tuatara sample from seven islands, two from Middle mercury, three from North Brother and 22 from Stephens Islands (which has the



largest population). To examine the phylogenetic relationships among tuatara from ten distinct islands, a maximum likelihood tree was constructed using complete mitochondrial genome sequences (Figure 1). We also obtained similar topology for the ML trees constructed using constrained and neutral sites (Figures S1 and S2 respectively – Supplementary material). For this purpose we used the diversities at synonymous sites (dS) and constrained sites (dC) respectively. The constrained sites consist of concatenated alignments of nonsynonymous sites of 12 protein-coding genes (gene NADH dehydrogenase subunit 5 is absent in tuatara²), rRNAs and tRNAs (see Methods).

Since tuatara are the only surviving member of the order Sphenodontia there is no outgroup species available to root the tree. This is because the divergence time between tuatara and its closest relative, the Squamata reptiles is ~220 million years^{3,4}. However we did use the genome sequence of a chameleon (*Chameleo chameleo* - a Squamata reptile) and reconstructed the ML tree using only the constrained sites. As shown in Figures S3 and S4 (supplementary material) the topology is similar to the tree shown in Figure 1.

Our samples are from a wide geographic range, including locations on off-shore islands in the North Island and Cook Strait of New Zealand. While the shortest distance between two islands was

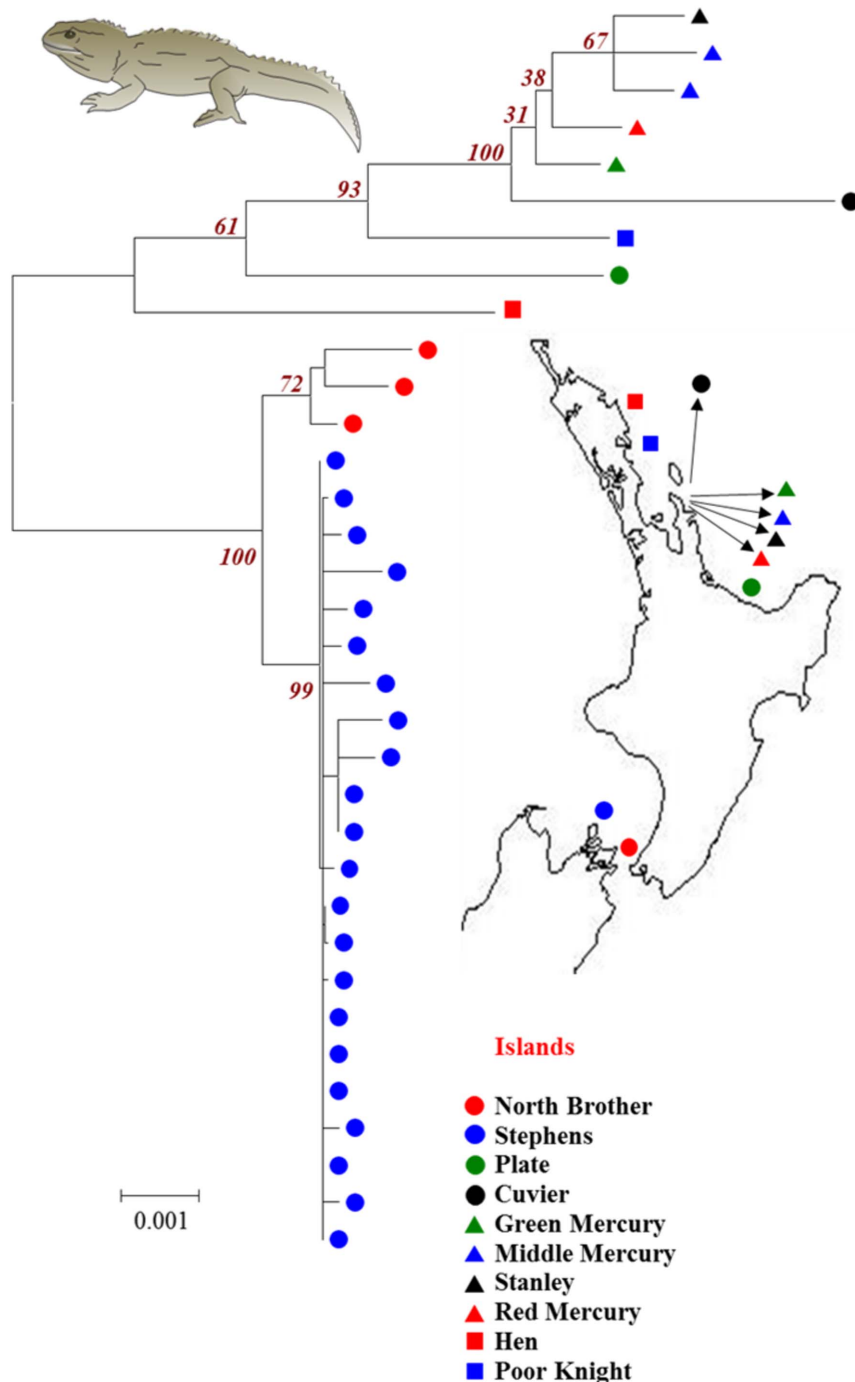


Figure 1 | A maximum likelihood tree showing the phylogenetic relationships among tuatara from ten offshore islands of New Zealand. The tree was constructed using the complete mitogenome sequences. The bootstrap support (1000 replications) for major nodes are indicated. The map shows the approximate locations of the samples used in the study. The map and the image of tuatara were manually drawn by us using the software Canvas (ACD systems international Inc.) and Adobe illustrator (Adobe Systems Inc.).



1.3 km (Green and Middle Mercury Islands) the longest was 631 km between Poor knight and Stephens Islands (Figure 1). We examined the amount of diversity at neutral and constrained sites that could be explained by the distance between the islands. We plotted pair-wise diversities (dC and dS) between tuatara from different islands against the geographical distance between them. Figure 2A shows positive relationships between geographic distance with dC (Spearman $\rho = 0.71$) and dS ($\rho = 0.92$). These were highly significant using Mantel's test ($P < 0.001$). However a closer examination revealed that the relationship of dS with geographic distance is approximately linear and that of dC is non-linear. While dS estimates varied between 0.001 to 0.06 substitutions per site (s/s), dC range was much smaller between 0.001–0.01 s/s. However in Figure 2A we overlaid the relative diversities of dC and dS for each pair-wise comparison on the Y and Z axes respectively. This revealed that dC values estimated for pairs of individuals from closely located islands (eg. Stanley and Middle Mercury) were much higher than their corresponding dS. In contrast dC observed for pairs of tuatara individuals living in distantly separated islands (eg. Cuvier and Stephen) were relatively less than their corresponding dS. To demonstrate this clearly we computed the ratio of dC/dS for all pair-wise comparisons and plotted them against the geographic distances (Figure 2B). This resulted in a negative correlation (Spearman $\rho = -0.82$) that was highly significant ($P = 0.001$). The declining trend of dC/dS could be seen clearly after rescaling the Y-axis to exclude the outlier in order to focus on other data points (Figure 2C) (The outlier was not removed and it was included for calculating correlation coefficient and for determining the statistical significance). The mean dC/dS ratio estimated for the comparison involving tuatara living in islands that were separated by <100 km was 0.35. In contrast, this ratio was only 0.13 for the pairs living in islands that are >500 km apart, which was 2.7 times smaller than that of the former. We also estimated the net diversities between populations by subtracting the intra-specific diversity from inter-specific diversities (see Methods). The correlations between the geographical distance and net diversities were also highly significant. The observed positive correlations between geographical distances and net-dC (Spearman $\rho = 0.56$) as well as between geographic distance and net-dS (Spearman $\rho = 0.91$) were highly significant ($P < 0.001$) using Mantel's test. A highly significant negative relationship (Spearman $\rho = -0.74$, $P = 0.001$) was observed between geographic distance and the ratio of net-diversities at constrained and neutral sites (dC/dS).

Discussion

The highly significant positive relationship between dS and geographic distance illustrated in this study suggests the possibility that neutral diversity increases proportionally with increasing distance between tuatara populations. We also examined the strength of the above relationship using the mitochondrial D-loop region, which was available for 58 individuals from the ten islands (2–11 individuals from each island) used in this study¹². The results of this analysis also revealed a significant positive correlation between the geographical distance and genetic diversity at D-loop (Spearman $\rho = 0.68$; $P = 0.001$) (Figure 3A). This suggests that the correlation analyses reported in this study using complete mitogenomes are robust.

We also estimated the fixation index (F_{ST}) between tuatara populations in order to examine the relative contributions of drift and migration (gene flow) using the D-loop data. The mean inter-population D-loop diversity was estimated to be 0.03. In contrast, the average intra-population diversity was only 0.005. The resulting F_{ST} was 0.85, which suggests a much smaller proportion of within population diversity. This suggests that between-(island)population diversities estimated in this study were not significantly affected by intra-population diversities. To further examine this we correlated pairwise F_{ST} estimated for populations from different islands with

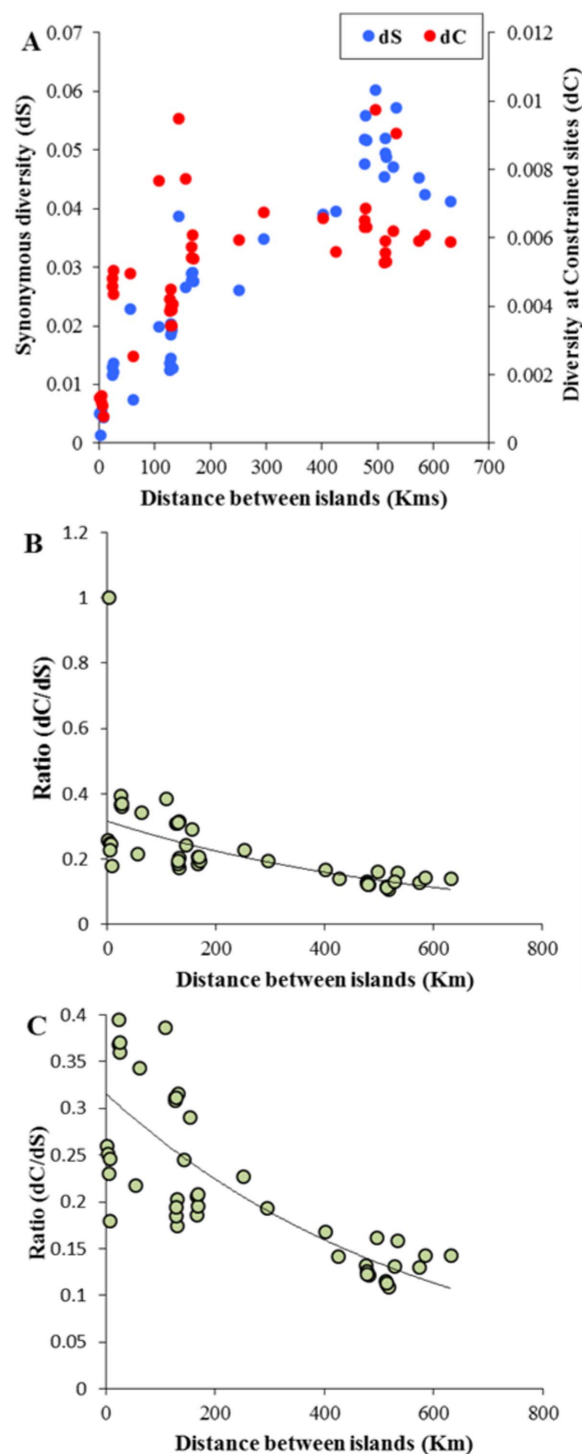


Figure 2 | (A) Relationship between geographic distance and diversities at synonymous (Spearman $\rho = 0.71$; $P < 0.001$) and constrained (Spearman $\rho = 0.92$; $P < 0.001$) sites including the nonsynonymous sites of 12 protein-coding genes, tRNAs and rRNAs. The relationships were highly significant based on Mantel's test (B) Correlation between geographic distance and the ratio of constrained- to synonymous sites (dC/dS) (Spearman $\rho = 0.82$; $P = 0.001$). (C) Same as Figure 2B but the outlier is not shown. (The outlier was not removed and it was included for calculating correlation coefficient and for determining the statistical significance. Y-axis was rescaled to exclude the outlier in order to focus on the other data points).

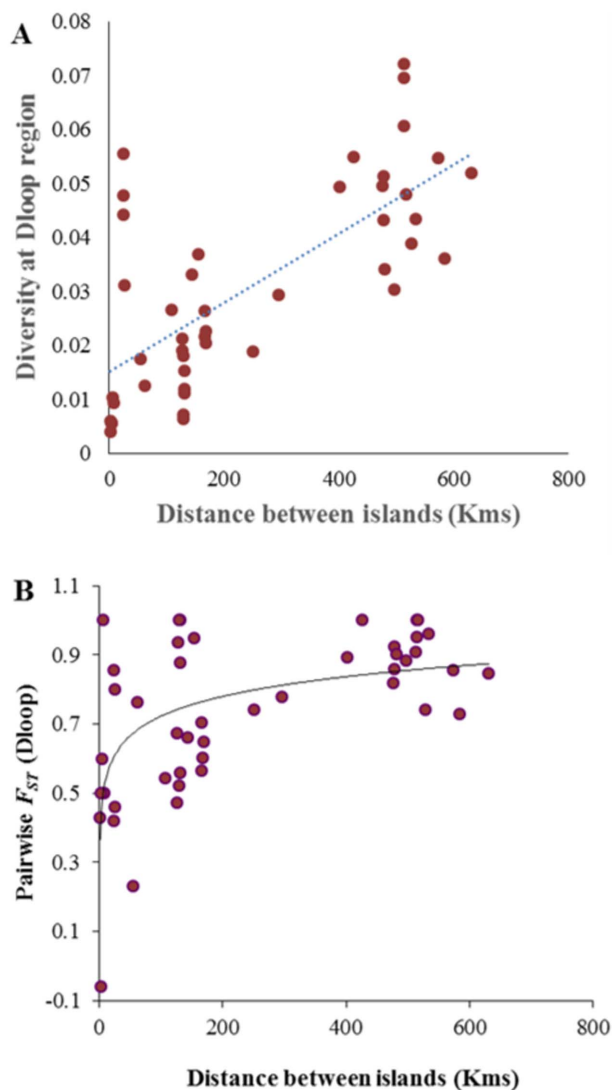


Figure 3 | (A) Relationship between geographic distance and diversity at the D-loop region of tuatara mitogenomes (Spearman $\rho = 0.68$; $P = 0.001$). (B) Correlation between geographic distance and F_{ST} estimated for populations from a pair of islands (Spearman $\rho = 0.54$; $P = 0.004$). The relationships were highly significant based on Mantel's test. Best fitting regression lines are shown.

the geographic distance between the islands. We observed a highly significant positive relationship (Spearman $\rho = 0.54$; $P = 0.004$) suggesting increase in population structure corresponds with the increase in the geographic distance between the island populations (Figure 3B). Since the mitogenome data from multiple individuals was not available for seven island populations we could not perform this analysis using the genome data. However the F_{ST} analysis from D-loop data imply that the relationship between genetic and geographic distances appear to be largely determined by genetic drift. This also suggests that migration rates of tuatara populations between islands might be minimal.

The positive relationship between geographic distance and neutral diversity suggests a direct relationship between geographic distance and time of divergence between the island populations. Further evidence for this prediction comes from our dC analysis. It is well known that at short timescales deleterious mutations are expected to segregate in populations due to the power of genetic drift. However they are selected against over time and are eventually eliminated from a population^{13,14}. Previous studies using intraspecific comparisons of

human mitochondrial genomes observed 5–10 times higher proportion of deleterious variations compared to inter-species divergence^{15,16}. In this study we observed a high dC/dS for the comparisons involving tuatara living in proximal islands and much lower dC/dS values for the pairs of individuals inhabiting distal islands. The different dC/dS ratios actually imply the difference in the divergence times between tuatara populations. The observed higher dN/dS ratio could also be due to the effects of population sizes as small populations harbour more slightly deleterious mutations. However tuatara population sizes of nine of the islands examined in this study are extremely small (<1000) except for that in the Stephens island (~30,000) (<http://www.doc.govt.nz/documents/science-and-technical/TSRP47.pdf>) and hence any significant effects of population size on our results is unlikely. The higher dC/dS ratios (1.0) observed might also suggest the possibility of positive selection. However it is well known that the effect of genetic drift is much higher in small populations, which reduces the efficacy of natural selection^{13,14}. Therefore the observed relationships cannot be explained by positive selection as the population sizes of tuatara living in most of the islands are small.

Studies of population genetics, evolution and ecology routinely examine the relationship between genetic and geographic distances in order to infer the demographic history of populations. However many studies use genetic data to estimate population genetic parameters without distinguishing whether the sequences are under selective constraints or evolve neutrally. Here we have shown that these two types of genomic regions behave quite differently. This suggests that future studies need to test their sequence data for neutrality to obtain unbiased relationship between genetic and geographic distances.

The temporal pattern of deleterious non-synonymous polymorphisms has been demonstrated by a number of previous studies^{15,17–19}. Here we identified (geographic) distance-dependent patterns of divergences using island populations of tuatara. Our results suggest that the geographic distance that separates two populations (or closely related species) of land vertebrates can be used to quantify the relative time of divergence between them particularly when the dispersal or migration rate is minimum.

Methods

Blood samples were collected from 34 tuatara individuals from ten islands around New Zealand (number of samples in parenthesis): Poor Knights (1), Cuvier (1), Stanley (1), Green Mercury (1), Plate (1), Hen (1), Red Mercury (1), Middle Mercury (2), North Brother (3) and Stephens Islands (22) (Fig 1). DNA was extracted from blood samples²⁰ and the entire mitochondrial genome of each tuatara individual was amplified in 16 overlapping fragments each approximately 1–2 kb in length. The sequencing was done on two independent PCRs. One PCR product was sequenced in the forward direction and the other one in reverse. In cases, in which there was a SNP in one direction but the second sequence was not long enough to cover that SNP, or showing a different SNP, then the third PCR was performed. The complete tuatara mitochondrial genomes were obtained using long-range PCR of 1–2 kb DNA fragments. The Long-Range PCR carried out in a 20 μ l volume containing 1X PCR buffer (Invitrogen), 50 mM MgCl₂ (Invitrogen), 1 mg/ml BSA (Invitrogen), 250 μ M mix dNTPs (BioLine), 0.50 μ M forward primer (Invitrogen), 0.50 μ M reverse primer (Invitrogen), 5 U Elongase Taq and 1 μ l DNA template. The PCR mixes were amplified using an iCycler™ Thermal cycler (BIO-RAD, USA). The amplification programme consisted of initial denaturing at 94°C for 4 min followed by 10 cycles consists of 94°C for 30 sec, 57.5°C (primer specific) for 30 sec, 8 min extension at 68°C and another 30 cycles consists of 94°C for 30 sec, 57.5°C (primer specific) for 30 sec and 68°C for 8 min. For the following 30 cycles the extension time was increased by 20 sec after cycle one. The final extension at 68°C for 4 min was followed. The PCR products were stored at 4°C for further analysis. The PCR products were subjected to electrophoresis in 1.5% agarose, stained with ethidium bromide (50 ng/ml) and visualized over ultra violet (UV) light.

The list of primers used are given in the Supplementary material (Tables S1–S3). For primer design and optimization the program primer 3.0 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) was used. For limiting the DNA amplification to the long fragments and avoiding the amplification of shorter sequences, particular PCR conditions were used. This resulted in competition between the templates in each cycle in favor of the long template from the preceding cycle. After purification each fragment was sequenced in both directions using forward and reverse primers using an ABI 3730 sequence analyzer. This was important to verify (any) nucleotide



variants. A third PCR was performed in the case of any inconsistencies. All methods were carried out in accordance with the approved guidelines of Griffith University, Nathan, Australia. All experimental protocols were approved by Griffith University, Nathan, Australia.

The complete mitochondrial genomes belonging to 34 tuatara individuals along with the reference (NC 004815) genome were aligned using the program *SeaView*²¹. The reference genome was used to identify the boundaries of protein-coding and noncoding sequences. To identify the most appropriate evolutionary models and other parameters, we used the program MODELTEST²² as implemented in MEGA²³. Based on Bayesian Information Criterion (BIC) we found the model Tamura-Nei + Gamma best described the data. Using this model a maximum likelihood tree was constructed using MEGA. To obtain statistical support for each node we used the bootstrap resampling procedure with 1000 replications.

To estimate pair-wise diversities at synonymous sites (dS) we used third codon positions (3182 bp) or four-fold degenerate sites (1161 bp). Although both produced almost identical results we used the former for all analysis due to its size. For constrained sites (dC) we concatenated nonsynonymous (first two codon positions) sites of protein-coding genes, all tRNAs and two rRNAs (10360 bp in total). To estimate genetic diversities within and between populations we used HKY + Gamma model for constrained and synonymous sites using the alpha values of 0.05 and 0.11 respectively. This model was chosen using the program MODELTEST and it corrects for multiple substitutions, accounts for base compositional bias as well as corrects for rate variation among sites. To estimate the net diversity between populations we used Nei and Li's method $d_A = d_{XY} - (d_X + d_Y)/2$, where d_{XY} is inter-population diversity and d_X and d_Y are intra-population diversities²⁴. From a previous study¹² we obtained 58 D-loop sequences of tuatara living in Green Mercury (3), Middle Mercury (2), Red Mercury (2), Stanley (3), Cuvier (4), Plate (6), Hen&Chickens (12), Poor Knights (11), Stephens (8) and North Brother islands (7). We excluded the flanking tRNAs and used only the D-loop region (928 bp) spanning between positions 95 to 1022 of the sequences reported by the previous study¹². To estimate diversity we used TN92 + Gamma model using a shape parameter of 0.05.

Geographic distances between individuals from pairs of islands were obtained from the online resource *DistanceFromTo* (<http://www.distancefromto.net/>). To test the significance of any correlation between geographical distance and genetic divergence we used the program *Isolation by distance*²⁵. Specifically we performed a Mantel test that is based on Spearman correlation coefficient, as some of the relationships reported here were nonlinear. However the strength of the relationship and the levels of statistical significance obtained using the Mantel test based on Pearson correlation coefficient (employed in the program *Genepop*²⁶) were similar. We have reported the correlation coefficients and significance for the power law (log-log) relationships. However the significance observed for linear-linear or log-linear relationships were also highly significant (at least $P < 0.007$). To estimate pairwise F_{ST} between tuatara populations we used the software *DNASP*²⁷.

- Cree, A. & Butler, D. *Tuatara recovery plan* (New Zealand Department of Conservation, Wellington, 1993).
- Rest, J. S. *et al.* Molecular systematics of primary reptilian lineages and the tuatara mitochondrial genome. *Mol Phylogenet Evol* **29**, 289–297 (2003).
- Benton, M. J. *Vertebrate Paleontology (2nd Edition)*. (Backwell Science, 2000).
- Carroll, R. L. *Vertebrate Paleontology and Evolution*. (W.H. Freeman & Co., 1988).
- Duncan, R. P., Blackburn, T. M. & Worthy, T. H. Prehistoric bird extinctions and human hunting. *Proc Biol Sci* **269**, 517–521, doi:10.1098/rspb.2001.1918 (2002).
- King, M. *The penguin history of New Zealand*. (Penguin Books, 2003).
- Wright, S. Isolation by Distance. *Genetics* **28**, 114–138 (1943).
- Li, J. Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104, doi:10.1126/science.1153717 (2008).
- Maes, G. E. & Volckaert, F. A. M. Clinal genetic variation and isolation by distance in the European eel *Anguilla anguilla* (L.). *Biol J Linn Soc* **77**, 509–521.
- Pistis, G. *et al.* High differentiation among eight villages in a secluded area of Sardinia revealed by genome-wide high density SNPs analysis. *PLoS one* **4**, e4654, doi:10.1371/journal.pone.0004654 (2009).
- Xing, J. *et al.* Genetic diversity in India and the inference of Eurasian population expansion. *Genome Biol* **11**, R113, doi:10.1186/gb-2010-11-11-r113 (2010).
- Hay, J. M., Sarre, S. D., Lambert, D. M., Allendorf, F. W. & Daugherty, C. H. Genetic diversity and taxonomy: a reassessment of species designation in tuatara (*Sphenodon*: Reptilia). *Conserv Genet* **11**, 1063–1081 (2010).

- Kimura, M. *The neutral theory of molecular evolution*. (Cambridge University press, 1983).
- Ohta, T. The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst* **23**, 263–286 (1992).
- Hasegawa, M., Cao, Y. & Yang, Z. Preponderance of slightly deleterious polymorphism in mitochondrial DNA: nonsynonymous/synonymous rate ratio is much higher within species than between species. *Mol Biol Evol* **15**, 1499–1505 (1998).
- Parsons, T. J. *et al.* A high observed substitution rate in the human mitochondrial DNA control region. *Nat Genet* **15**, 363–368, doi:10.1038/ng0497-363 (1997).
- Ho, S. Y., Phillips, M. J., Cooper, A. & Drummond, A. J. Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Mol Biol Evol* **22**, 1561–1568, doi:10.1093/molbev/msi145 (2005).
- Subramanian, S. Temporal trails of natural selection in human mitogenomes. *Mol Biol Evol* **26**, 715–717, doi:10.1093/molbev/msp005 (2009).
- Subramanian, S. *et al.* High mitogenomic evolutionary rates and time dependency. *Trends Genet* **25**, 482–486, doi:10.1016/j.tig.2009.09.005 (2009).
- Sambrook, J. & Russell, D. W. *Molecular cloning: a laboratory manual*. (Cold Spring Harbor Laboratory).
- Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol Biol Evol* **27**, 221–224, doi:10.1093/molbev/msp259 (2010).
- Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
- Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**, 2731–2739, doi:10.1093/molbev/msr121 (2011).
- Nei, M. & Kumar, S. *Molecular evolution and phylogenetics*. (Oxford University Press, 2000).
- Jensen, J. L., Bohonak, A. J. & Kelley, S. T. Isolation by distance, web service. *BMC Genet* **6**, 13, doi:10.1186/1471-2156-6-13 (2005).
- Raymond, M. & Rousset, F. Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism. *J Hered* **86**, 248–249 (1995).
- Librado, P. & Rozas, J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452, doi:10.1093/bioinformatics/btp187 (2009).

Acknowledgments

We wish to thank C.H. Daugherty and N. Nelson for providing us with samples. This research project was supported by a scholarship to E.M. from the Allan Wilson Centre for Molecular Ecology and Evolution. The project was also supported by Griffith University.

Author contributions

S.S. and D.L. conceived the project. E.M. performed sequencing and all other laboratory analyses. C.D.M. contributed to sample collection and laboratory analyses. S.S. conducted data analysis and hypothesis testing. S.S. wrote the paper with inputs from all authors. D.L. and C.D.M. edited the paper.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Subramanian, S., Mohandesan, E., Millar, C.D. & Lambert, D.M. Distance-dependent patterns of molecular divergences in tuatara mitogenomes. *Sci. Rep.* **5**, 8703; DOI:10.1038/srep08703 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>