MSc Research Project Thesis, Biology, Leiden University

Genetic Disparity increases with Isolation in Insular Populations of the Introduced Polynesian Rat, *Rattus exulans*

Grace Saville

Student Number: 2640783

Understanding Evolution, Naturalis Biodiversity Center

Supervised by Dr Rutger Vos

20/09/2021 – 23/05/2022, 40 EC

Contact: gracei.saville@gmail.com

Formatted following PeerJ guidelines

Genetic Disparity increases with Isolation in Insular Populations of the Introduced Polynesian Rat, *Rattus exulans*

Grace I. Saville[12]

[1] Understanding Evolution, Naturalis Biodiversity Center, Leiden, South Holland, Netherlands
[2] Biology, Leiden University, Leiden, South Holland, Netherlands

Corresponding Author:
Grace Saville
Email address: gracei.saville@gmail.com

Abstract

Background. The Polynesian rat, *Rattus exulans*, accompanied travellers across the Pacific ocean during the period of the Austronesian expansion, colonising these islands. Examining the population genetics of Polynesian rats from different islands along the theorised ocean route can add to the evidence of the means and route of both rats and humans, and can reveal how island ecology has had an effect on the rat populations. In particular, that there is a correlation between geographic distance and the genetic differences seen in specimens from island pairs, and that homozygosity and therefore inbreeding will increase with each colonised island further away from the southeast Asian mainland.
Methods. Using a dataset of 298 short nucleotide polymorphisms (SNPs) and 370 rat specimens from 25 islands, I conducted a NeighborNet haplotype network reconstruction, Mantel test of genetic and geographic distance, Homozygosity examination, Fixation index (F_{ST}) computation, and several linear regression models.
Results. The results support the hypotheses: populations further away from each other have more genetic differences and those closer have fewer (Mantel observation of 0.499, p-value of 0.001), and those further from the mainland are more inbred (linear model p-value of 0.002). However distance of an island does not explain all genetic variation or inbreeding, and this result should only be used as a simplification or generalisation. Other factors such as biased sampling, ongoing human trade between islands or the presence of eradication programmes can have an effect on the Polynesian rat gene pool.

Introduction

If you took Disney Pixar's Ratatouille and crossed it with Moana, perhaps you would get an approximation of the life of a Polynesian rat during the time of human exploration of the Pacific islands. The Polynesian rat, *Rattus exulans*, also known as kiore or the Pacific rat, lives up to its name through its presence in most of the islands on the ocean between southeast Asia and South America. The Polynesian rat inhabits islands that range in size from below 1km² to over 785,000km² (New Guinea) and latitudes from 0˚ to 50.5˚ (Auckland Islands, New Zealand) (van der Geer, 2018). The presence of *R. exulans* on these islands is thought to

be facilitated by humans during the Austronesian expansion. The Austronesian expansion was the step-wise radiation of humans from one island to the next over the course of generations, colonising many as they went. The route humans took traces from Taiwan (Mirabal et al., 2013), through the Bismarck Archipelago to Vanuatu, Samoa and Fiji, the Cook and Society Islands, followed by the Marquesas and reaching its easternmost point at Rapa Nui (Easter Island) (Ioannidis et al., 2021) and southernmost point at New Zealand (Aotearoa). Evidence of this approximate order of island hopping has been obtained through genomic, linguistic and archaeological studies (for example, see Lum & Cann, 1998; Carson et al., 2013; Matisoo-Smith, 2015). The Polynesian rat origin point is suggested to have been not Taiwan, but the island of Flores in Indonesia, based on genomic studies (Thomson et al., 2014), although there is currently no (sub)fossil evidence to support this (van der Geer, 2020). This order of rat and human island colonisation is highly simplified, because there was ongoing interaction between many islands that has led to unclear evidence of the "finer-scale" timing and peoples who first arrived (West et al., 2017). Human populations on remote islands maintained contact with other islands, seen for example through the trade of adzes, for many centuries (Collerson & Weisler, 2007; Weisler et al., 2016), therefore the estimates made on the timing of initial human colonisation are conflicting, some 1000 years apart (Wilmshurst et al., 2008, 2011).

Because rats and humans made this expansion concurrently, the Polynesian rat can be used as a proxy for the movement of humans and vice versa (Matisoo-Smith, 1994; Matisoo-Smith et al., 1998), although the evolutionary changes that may have occurred in *R. exulans* during this process are important in their own right. The Polynesian rat makes a good proxy because they were among the earliest species to be transported with Polynesian explorers (given subfossil presence in the first archaeological layers of historical sites (Barnes, Matisoo-Smith & Hunt, 2006)), they have a broad distribution, many extant populations are still present on those islands today, and they are reproductively isolated from other introduced rats such as *R. norvegicus* and *R. rattus* (Matisoo-Smith & Robins, 2009). The success of their colonisations and current range could also be attributed to being a generalist species, capable of surviving in a range of habitats, climates and diets (Campbell et al., 1984), particularly without competition or predation (Athens, 2009). Other proxies for human movement studied have included pigs (Larson et al., 2007) and chickens (Matisoo-Smith & Robins, 2004), the paper mulberry plant (Matisoo-Smith, 2015), stone adzes (Collerson & Weisler, 2007; Weisler et al., 2016), and kumara (sweet potato), the latter originally a South-American tuber giving evidence towards a voyage route beyond Rapa Nui (Roullier et al., 2013). Trade occurred between the islands and colonisers brought familiar plants and animals, such as those mentioned, with them (Kirch, 2000). In the case of the Polynesian rat, they may have been brought on voyages as a food source and intentionally allowed to colonise islands found (Roberts, 1991; Matisoo-Smith & Robins, 2004).

Island biogeographical and evolutionary processes also play a part in the gradual Pacific island colonisation process and the evolution of the Polynesian rat. "Island syndrome" refers to a collection of attributes that characterise species living isolated on islands for a period of time. These attributes include gigantism (van der Geer, 2018), dwarfism, loss of wings, colour changes (for example, relaxed sexual dimorphism colouring or increased melanism), reduced brain size (animal species can be slower and less alert), increased docility, and reduced territorialism (Adler & Levins, 1994). These traits can also be linked to domestication syndrome, which has similar effects (Wilkins, Wrangham & Fitch, 2014; Geiger, Sánchez-

Villagra & Lindholm, 2018). Research by van der Geer (2018, 2019, 2020) suggests that *R. exulans* displays island syndrome (and thereby also some characteristics of domestication syndrome) through increased coat colour polymorphisms and increased body size in remote eastern islands and in New Zealand. Traits like these can be caused by reduced competition and predation (Adler & Levins, 1994). Species that show the characteristics of island syndrome tend to be sensitive to introduced predators as they can be ecologically naïve, however in this case *R. exulans* themselves are the introduced predator to many islands (Harper & Bunbury, 2015), for example New Zealand (Gibbs, 2009). Island biogeography has also affected *R. exulans*, particularly the founder effect, similarly to humans, as "the genetic composition of each remote island group is dominated by the contribution of its founders" (Ioannidis et al., 2021). Other effects that may have played a part in the Polynesian rat movement include colonisation ability given the distance to the mainland or nearby islands, the rate of extinction being lower on larger islands (because of more space, different habitats, and less chance of an all-encompassing extinction event), the rate of extinction being lower on closer islands (more colonisers) (MacArthur & Wilson, 1967), length of isolation, climate, other species and resources present. This project aims to add to the evidence of the means and route of Polynesian rat dispersal through the Pacific and the effect island ecology has had on them, specifically by considering the relationship between the distance between islands or the distance to the mainland and the level of genetic variation or inbreeding present in insular populations of *R. exulans*. Much of the current literature on the movement of the Polynesian rat (and humans, both directly and by proxy) focusses on unravelling the finer details of island route and colonisation, such as the specific time of arrival for each island, however this project has an alternative, broader goal of building on the fundamentals of island biogeography processes and providing a model or system of analysis for other island hopping species, particularly those without the benefit of suitable proxies.

Preceding this research project, 478 *R. exulans* specimens were collected from 25 different islands (Fig. 1) and the southeast Asian mainland. Eight were deep-sequenced in order to asses which biallelic single nucleotide polymorphism (SNP) loci were highly variable and therefore suitable for analysis on genetic variation. Following a run through the KASP pipeline (Semagn et al., 2014), 317 SNP loci were returned. The resulting table of the 317 SNP loci of the 478 *R. exulans* specimens is the raw dataset used in this research project (available at https://github.com/gsaville/kiore-project.git). This dataset was further examined for outliers and missing data, as described in the methods section below.

To evaluate the relationship between genetic distance and geographic distance, homozygosity, and distance from the mainland, analyses were conducted by running multiple tests on the SNP dataset (cleaned down to 370 rat specimens and 298 SNP loci), including a NeighborNet haplotype network, Mantel tests, Homozygosity test, Fixation index (F_{ST}) computation, and linear regression modelling using variables such as island distance from the mainland, island area (a proxy for ecological diversity and population size), homozygosity level, F_{ST} "distance", pairwise geographic distances between islands, and pairwise genetic distances between *R. exulans* specimens on those same islands.

The research question guiding this research project is "How have island evolutionary processes and human impacts affected the expansion and adaptations of the Polynesian rat, *Rattus exulans*, over the Southwest Pacific?". To assist in answering this question, three sub-questions have been raised to guide the research, as follows; i) "Which *R. exulans* island

populations demonstrate the closest genetic relationships?", ii) "What is the correlation between geographical distances of insular *Rattus exulans* populations and genetic distances?" and iii) "How prevalent is increased homozygosity in Pacific *Rattus exulans* populations successively further from the mainland populations, and is inbreeding correlated with island size?". The results are expected to reveal that neighbouring islands (which are also presumably islands that humans colonised closely in time to each other, if at all) harbour Polynesian rat populations that are the closest genetically. Regarding homozygosity, it is expected that it will increase with each island that is slightly further away from the mainland population in southeast Asia. This is because with each successive colonisation, presumably assisted by humans, the founder effect is likely to occur, particularly on islands without gene flow.

Materials & Methods

Data Clean-up: Missing and Monomorphic data removal.

Within RStudio, I replaced equivalent SNPs in the dataset (available at https://github.com/gsaville/kiore-project.git) with the same characters (for example, "A:T" and "T:A" both became "A:T"). I checked for invariant/monomorphic columns, i.e. loci with no variation, by calculating the number of distinct values per locus (the possibilities being homozygous base pairs, "A:T", "A:C", "A:G", "C:T", "C:G", "T:G" and "?"), to look for loci that had only one or two distinct values ("?" included). I found no monomorphic loci, which was in line with how the SNP loci were originally selected for their variability. The removal of monomorphic loci is important because if every specimen in the dataset has the same SNP at a particular locus, the SNP would not contribute to assessing the genetic differences between the different Polynesian rats.

I removed specimen rows and locus columns with high levels of missing data by calculating the percentage of question marks per row and column and removing those above a certain threshold. The threshold for locus columns was 60%, so those above 60% missing data were removed, and the threshold for the specimen rows was 56%. The latter was arrived at empirically as with at least 44% SNP completeness per row, the whole data frame was over 90% complete (90.277%).

Data Review: Hardy-Weinberg Analysis.

The Hardy-Weinberg analysis was conducted to check for loci within each island population (including the mainland population, which is implied throughout) that are out of Hardy-Weinberg equilibrium (Weinberg, 1908; Hardy, 1908; Edwards, 2008). Hardy-Weinberg equilibrium is when the alleles in a population are in balanced proportions, theoretically assuming there is assortative mating but no natural selection nor genetic drift (Wigginton, Cutler & Abecasis, 2005). Hardy-Weinberg disequilibrium can occur in the loci of a population when there is, for example, genetic drift, inbreeding, or potentially a methodical issue with the dataset. The dataset was converted to a CONVERT (Glaubitz, 2004) formatted text file by replacing the adenine (A), thymine (T), guanine (G) and cytosine (C) symbols with "1", "2", "3" and "4" respectively and partitioning the specimens into their respective populations. Using PGDSpider (Lischer & Excoffier, 2012) I exported the CONVERT data as an Arlequin (Excoffier & Lischer, 2010) input file. Within Arlequin (Excoffier & Lischer, 2010) I ran an exact test of Hardy-Weinberg equilibrium, with an allowed missing level per site of 0.6. Since there were many comparisons within the analysis, I did a Holm's Sequential Bonferroni adjustment on the *p*-values before examining those that were significant. There

was no more than one population per locus with a significant *p*-value indicating disequilibrium (with one exception, locus 41 with two populations), therefore I accredited that to either data error or bias, an outlier, or interbreeding with another population. I did not remove any loci because there was no clear trend of disequilibrium in any one locus. I did however notice that there were 17 loci with issues in the Kayangel population, which indicated there may be an issue with those specimens, not with the loci. Using a preliminary "pre-clean-up" NeighborNet (Bryant & Moulton, 2004) haplotype network I identified a cluster of six specimens from Kayangel that I decided to remove since I believe them outliers, perhaps due to a data collection error.

Data Review: STRUCTURE Analysis.

The purpose of STRUCTURE analyses in general is to derive an estimate of the number of hypothetical ancestral populations that gave rise to the observed patterns of diversity. Here, I used this method as way to detect outliers: if a specimen is estimated to be structured by ancestral components at odds with the prevalent structure in the labelled source then it may be an anomaly also. The analysis began too with converting the PGDSpider (Lischer & Excoffier, 2012) input file to one compatible with the STRUCTURE program (Pritchard, Stephens & Donnelly, 2000). The resulting STRUCTURE input file was used to run the test with 50,000 MCMC replicates and a burn-in of 5,000.

To assist in parsing the output, I used Structure Harvester (Earl & VonHoldt, 2012). It identified $\Delta K = 13$ (under the Evanno method (Evanno, Regnaut & Goudet, 2005)) as the most likely number of ancestral populations. The results identified one specimen from Kamaka and six from Kayangel (as stated above) as genetically different from their comrades. These were removed (if not already).

The specimens from Kamaka (population nine, Fig. 2) have 93.7% of the genetic data placed in the dark green cluster, however one specimen in that cluster (dark purple bar) shares a genetic structure more closely with specimens from New Guinea (pop. 17, 98.9% dark purple), Normanby Island (pop. 18, 96.4% dark purple), Malenge (pop. 13, 68.7% dark purple), New Britain (pop. 16, 67% dark purple), and several other islands from the same region (Table S1), indicating that the specimen was potentially labelled wrong or mixed up during a step such as sequencing. That particular specimen (labelled Kamaka 009) is missing 56% of the SNP data, compared to the other Kamaka specimens which are missing 1-5% SNP data. Population ten (Kayangel) mostly has a distinct bright purple colour (69.9%) which it does not share with any other populations, however there are six specimens that are apparently close to specimens from Sulawesi (pop. 23, 90.1% blue), the mainland (pop. 12, 58.7% blue), Halmahera (pop. 5, 35.8% blue), and several other nearby islands (Table S1). The Hardy-Weinberg analysis also identified some issues with the loci in the Kayangel population, which lead me to believe that the specimens in dark blue may be errors, if not an unexplained biological irregularity. The population Rimatuu (number 21, 85.7% cyan) also has a unique colour, with two specimens labelled a peach colour as being more closely related to population 20 (Reiono), however this is not an issue because in some STRUCTURE runs Rimatuu and Reiono are classed as one population, and the islands are only approximately 1600m apart.

NeighborNet Haplotype Network.
To create a NeighborNet (Bryant & Moulton, 2004) haplotype network and genetic pairwise distance matrix with the program SplitsTree (Huson & Bryant, 2006), I created a Phylip file (Felsenstein, 2005) that could be read by SplitsTree. First I replaced each SNP pair with a symbol from the IUPAC Ambiguity code (Johnson, 2010) as in Table 1, then used the R function dat2phylip() from the phylotools package (https://github.com/helixcn/phylotools).

The substitution model settings selected in SplitsTree were HKY85 (Hasegawa, Kishino & Yano, 1985) distances with empirical base frequencies. I exported the NeighborNet (Bryant & Moulton, 2004) tree diagram as an image and exported the distance matrix as a tab-delimited text file for use in Mantel testing (Mantel, 1967) later.

Pairwise Distances and Mantel test.
Before conducting the Mantel test (Mantel, 1967), I verified the spatial coordinates by mapping. I noticed that Tahanea was missing a longitude negative sign, which I replaced. Additionally, Reiono and Honuea had the same coordinates, for which I looked up the correct ones.

I created a pairwise geographic distance matrix between the islands by using the function distm() in the geosphere package (https://github.com/rspatial/geosphere), which uses latitude and longitude coordinates to calculate the great-circle distance (in other words, distance in a straight line over a globe) between the two points. With this matrix and the pairwise genetic distance matrix from SplitsTree (Huson & Bryant, 2006), I ran the Mantel test (Mantel, 1967) with 999 replicates.

I followed up the Mantel test (Mantel, 1967) with a linear model comparing genetic distance between islands against geographic distance, to test if there is indeed a linear relationship as hypothesised. The distribution of the residuals in the simple model “genetic distance ~ geographic distance” was positively skewed, and there was non-constant variance, neither of which could be remedied with a transformation, so I conducted a Generalised Linear Model (GLM) with the “identity” link function. Usually the gamma family uses the “log” or “inverse” link function, however when testing these they gave unusual coefficients (for example, intercepts of seven or negative two when approximately 0.1 was expected).

Homozygosity Testing.
To test if the level of homozygosity per island population increases as distance from the mainland increases, I replaced each SNP combination with “O” or “E” for homozygous and heterozygous SNPs. I sorted the specimens into their respective islands and calculated the total SNPs, homozygous SNPs, heterozygous SNPs and missing values per island, and the percentage of homozygosity. To the data, I added distance from the mainland and island area, the latter from island area data supplied by van der Geer (pers. comm., available at https://github.com/gsaville/kiore-project.git). The distances I calculated with the geosphere package, using the single Cambodian specimen as the mainland reference point.

I constructed three linear models; homozygosity ~ distance, homozygosity ~ area and combining both distance and area. The distance model had three outliers causing non-normality of the residuals and heteroscedasticity, which I decided to remove. The outlying points were the mainland, Normanby Island and New Guinea. I decided to remove the mainland point because it had high leverage and a value of zero, yet is not valuable since the

mainland is the reference point. Conducting tests without Normanby Island and New Guinea present (which are approximately 16.6km apart at the narrowest) saw an improvement in normality, which makes the model more reliable, however I do not think they are true outliers outside of the dataset, so the results from this model will come with this caveat. Additionally, New Guinea and Normanby island are based off one and three specimens respectively, and are missing more than half of the SNPs, which means the points may not represent their populations accurately.

The homozygosity ~ area linear model outlier testing also identified the points of New Guinea and Normanby island as a potential issue, and the model appeared more reliable without them (closer to residual normality and more equally distributed residual variance visible in the plots). In the homozygosity ~ area scatterplot there was also a large cluster of points close to the y axis (smaller islands), so I decided to log base 10 transform the area variable. This decision was supported by the boxcox plot, the fact that area is a squared measurement, and by the island biogeography concept that populations are less vulnerable the larger in area an island is (MacArthur & Wilson, 1967).

A Pearson's product-moment correlation test indicated that the island area and island distance from the mainland variables are correlated at -0.667 with a *p*-value of 0.0003, where zero is no correlation and -1 and +1 perfect correlation. In other words, bigger islands are nearer to the mainland coast, small oceanic islands are further away. Therefore I proceeded with caution when building the homozygosity ~ distance + area model, because the correlating explanatory variables affect the model outcome, despite collinearity tests indicating otherwise (condition number = 7.687 (10-30 indicates presence, issue if >30) and VIF = 1.727 (issue if >5)). To help compensate for this issue I conducted a GLM.

Fixation Index (F_{ST}) and Mantel test.

I used Arlequin (Excoffier & Lischer, 2010) to compute F_{ST} values on the clean dataset at a population level, therefore converted the data to an *.arp input file for Arlequin with PGDSpider (Lischer & Excoffier, 2012) the same as described during the process of the Hardy-Weinberg analysis. The settings in Arlequin (Excoffier & Lischer, 2010) for the F_{ST} test were an allowed missing level per site: 0.6, number of permutations: 1000 and significance level: 0.05. The resulting F_{ST} pairwise matrix was wrangled in RStudio and compared with an island pairwise geographic distance matrix once again created using latitude and longitude point and the geosphere R package (https://github.com/rspatial/geosphere). I produced a correlogram plot and built a linear model on this data. Within the linear model I investigated in particular the points with extreme residuals, and therefore an higher or lower F_{ST} than estimated by the model at any given geographic distance. Outlier testing during the diagnostics on the linear model identified New Guinea to mainland and Normanby Island to mainland as an issue, and they are the only two negative F_{ST} values. These negative values could indicate that the sample sizes are too small (Gerlach et al., 2010) or the populations have more variation within than between the island pair, which would not be surprising given that New Guinea is the largest island in this study and Normanby island is immediately off New Guinea's coast, and the mainland is not an island. I removed these two points when adjusting the model.

I conducted a Mantel test (Mantel, 1967) here also, between the F_{ST} "distance" and the geographic distance between island populations (as opposed to island specimens). I used the

geosphere package (https://github.com/rspatial/geosphere) to create an island geographic distance matrix to be compared against the F_{ST} matrix produced in SplitsTree (Huson & Bryant, 2006). The test was once again run with 999 permutations.

Results

NeighborNet Haplotype Network

In the network (Fig. 3), specimens from or closer to the southeast asian region appear to have longer lines therefore exhibit more variation than the specimens on the right hand side of the diagram (for example, the New Zealand islands), which appear to be densely clustered together with short lines, indicating less variation within and between the closest populations. New Guinea (orange) has particularly long edges. The mainalnd specimens appear to be closest to Sulawesi (red), and while many of the specimens from the southeast Asian region appear mixed with the other specimens from the same region, the Kayangel (dark yellow) specimens form a tight group. Late Island (pale pink) and Wake Island (salmon) specimens act as the bridge between the southeast Asian specimens and the remote island specimens. The remote island specimens, such as Aotea (royal blue) and Tahanea (pale green), cluster together in a similar pattern to the Kayangel specimens.

Pairwise Distances and Mantel test

Geographic distance tends to increase with genetic distance (Table 2), however there are some exceptions, such as Rimatuu and Rakiura being further away in kilometres from each other than Motukawanui and Rakiura, yet have a specimen pair that are closer genetically. Rimatuu and Reiono are the closest geographically yet have a specimen pair that are the eleventh closest genetically. The observed correlation between genetic distance and geographic distance from the Mantel test (Mantel, 1967) was 0.499 (with a simulated *p*-value of 0.001 based on 999 permutations). Zero suggests no correlation, i.e. genetic distance is not correlated with island distance, while negative one or positive one suggests a strong correlation, meaning closer islands would tend to have genetically closer populations. The result suggests that there is a significant positive correlation between genetic distance and geographic distance in our sample (and the null hypothesis of no correlation is rejected). The variance explained (R^2) for this observation is 24.876%. For comparison and reinforcement of this result, I also checked the results of a Pearson's product-moment correlation, which gave a correlation coefficient of 0.467 (*p*-value < 2.2e-16, $R^2 = 0.218$).

Generalised Linear Model.

The intercept and the slope of the genetic distance ~ geographic distance GLM was 9.434e-02 and 2.013e-05 respectively, both significant with an alpha of 0.05 ($p < 2e\text{-}16$), indicating a significant positive correlation. The calculated R^2 is 0.258, meaning 25.8% of the variance can be explained by the model, which is approximately the same as the R^2 computed in the Mantel test (and Pearson correlation). The distribution of the residuals in the GLM appears skewed, as was seen in the test model diagnostics (Min -1.742, 1Q -0.532, Median -0.22, 3Q 0.147, Max 3.028)

Homozygosity Testing.

The most inbred island population is Honea, with 96.03% homozygosity, while the least inbred is New Guinea at 72.73% homozygosity, which is slightly lower than the mainland level (Table 3). The mean is 89.38%, median is 91.5%, and the range is 23.3. The Kayangel Polynesian rat population has 94.69% homozygosity in this sample, which is unusually high

for how close it is to the mainland. Doubtful Sound, Luzon, Great Mercury Island and Normanby Island are only based on one specimen, and New Guinea and the mainland only on three and five specimens respectively (Table 3). Of these, New Guinea and Normanby Island are the only low-specimen samples without 100% complete SNP data, both missing more than half of the SNPs.

Linear Models.

In the Homozygosity ~ Distance from Mainland linear model, the intercept estimate is at 86.95, (p < 2e-16) and the Distance from Mainland coefficient is 5.221e-04, with a significant p-value of 0.002 when alpha is 0.05. The multiple R^2 for this model is 0.363, meaning only 36% of the variance of the residuals can be explained by this model.

For the Homozygosity ~ Island Area (log10) model, there is a significant negative correlation between the percentage of homozygosity and the area of the islands, so generally the smaller the island the greater the homozygosity. The intercept estimate is at 92.973 (p < 2e-16), and Island Area (log10) estimate at -0.917, also with a significant p-value of 0.006. This model explains slightly less of the residual variance than the distance from the mainland model, the R^2 being 0.306.

When both Distance from Mainland and Island Area (log10) variables are added to the model, the coefficients are 3.642e-04 for Distance from Mainland, and -0.199 for Island Area (log10), with p-values of 0.082 and 0.243 respectively (Fig. 4). The R^2 is slightly higher than the models with only a single variable, at 0.406. The spread of residuals for this model were Min -4.805, 1Q -2.046, Median -0.132, 3Q 2.058, and Max 4.743 identifying the residuals as close to normal distribution.

Fixation Index (F_{ST}) and Mantel test.

The F_{ST} value expresses the amount of genetic differentiation between two (sub)populations. An F_{ST} of one means the two compared populations have unique gene pools that do not share alleles, and the genetic variation within them is explained by the population themselves. An F_{ST} of zero means that the gene pools of both populations are blended as a result of gene flow, and therefore there is less genetic variation between the populations. Values under 0.15 indicate moderate differentiation between the populations, while values above 0.25 indicate particularly great differentiation (Wright, 1978). Negative values may be due to low data and can be treated as zero (Gerlach et al., 2010). The island pairs with the highest F_{ST} seen in Figure 5 are Honuea – Doubtful Sound (0.839), Honuea – New Guinea (0.830), Kayangel – Honuea (0.819), and Kayangel – Tahanea (0.803). Among the lowest F_{ST} values are combinations of the mainland, Borneo, Sulawesi, Halmahera and New Britain. Exact figures are reported in Table S2. The average F_{ST} (not including negative values) is 0.469, and the range is from 0.004 to 0.839, therefore on average 46.89% genetic variation between the groups. The average and maximum F_{ST} of the dataset is higher than expected for within-species analysis, given Wright's suggestion that an F_{ST} over 0.25 indicates extreme differentiation (Wright, 1978), however this estimate could be pulled upwards by higher levels of homozygosity on isolated islands. For example, the Honuea – Doubtful Sound pair (0.839) has a much higher F_{ST} than Honuea – Rakiura (0.463) yet similar distance. Although, the Doubtful Sound population in the dataset is based off one specimen, which reduces the confidence in this particular result.

The Mantel test (Mantel, 1967) between the F_{ST} matrix and pairwise geographic distance matrix gave the observed value of 0.377 (with a simulated *p*-value of 0.001 based on 999 permutations), where zero indicates no correlation and one (or negative one) indicates a strong correlation. This result rejects the null hypothesis that there is no correlation between F_{ST} values and geographic distance, in favour of the alternative hypothesis that the absolute value of the observed correlation is greater than zero. In other words, in this sample, islands which are closer together tend to have *R. exulans* populations with more shared genetics, while islands which are further away from each other tend to have rat populations which are more genetically isolated. The variance explained for this test was 14.193%.

Linear Model.

The model built here was the geographic distance between an island pair against the F_{ST} value of that pair. There is an even spread of residuals around zero, which is close to the median (Min -3.087, 1Q -0.816, Median -0.03, 3Q 0.884, Max 3.061). The intercept estimate is significant ($p < 2e\text{-}16$) at 0.37. The geographic distance coefficient is estimated at 1.889e-05, therefore only a slight slope, and a significant *p*-value of 4.96e-14. The multiple R^2 for this model is 0.175, therefore only approximately 17% of the variance of F_{ST} distance can be explained by geographic distance.

Island pairs with the most extreme negative residuals are Mainland – Malenge (-0.412 away from the mean, F_{ST} 0.004), Normanby Island – Borneo (-0.406, F_{ST} 0.037), Mainland – Halmahera (-0.376, F_{ST} 0.048), Normanby island – Halmahera (-0.368, F_{ST} 0.055), and New Britain – Borneo (-0.364, F_{ST} 0.075). In other words, these pairs demonstrate a closer relationship than would otherwise be expected based on the mean F_{ST} at their respective geographic distances from each other. Figure 6 shows that even though the distance is far between these islands around southeast Asia, the F_{ST} is low (blue lines). Further afield from the Sundaland and New Guinea regions, island pairs with particularly negative residuals include Normaby Island – Late Island (-0.278, F_{ST} 0.164), Wake Island – Late Island (-0.249, F_{ST} 0.21), Rakiura – Late Island (-0.206, F_{ST} 0.23), and mainland – Great Mercury Island (-0.202, F_{ST} 0.343). It is also visible in Figure 6 that some islands have rat populations that are more closely related to islands along the theorised colonisation route than those earlier in the route, for example the islands of New Zealand such as Rakiura, share lower F_{ST} values with Late Island (Fig. 1, F_{ST} 0.23) than with Borneo (F_{ST} 0.387). Kayangel (Fig. 1) is not visible in Figure 6 because it is the only island population where none of its F_{ST} value island pairs are less than the linear regression line.

Discussion

Addressing the Research Questions

What is the correlation between the geographical distances of insular R. exulans populations and the genetic distances? Which R. exulans island populations demonstrate the closest genetic relationships?

There is a significant positive correlation between the distance (in kilometres) between two islands, and the genetic distance between the specimens on those islands. In other words, generally islands close in great-circle distance are more closely related, and for islands that are further away from each other geographically, the specimens also are less similar genetically. This statement however, comes with a caveat. The differences in genetic distance cannot be entirely explained by how far away the islands are from each other, geographic distance only accounting for approximately 25% of the variance as can be seen in the pairwise

distance GLM, and 17% in the F_{ST}. Some of this variance is caused by the more genetically similar populations in southeast Asian and Australasian regions, specifically between the mainland, Borneo, Sulawesi, Malenge, Halmahera, New Guinea, New Britain and Normanby Island (blue lines, Fig. 6) despite their distance apart. These islands stand earlier in the theorised colonisation route, and the principle of island biogeography (MacArthur & Wilson, 1967) indicates that they are more likely to be re-colonised, either by the mainland or surrounding islands, therefore it is likely there is ongoing or recent gene flow, especially given that since those islands have had established rat populations for longer, the chances of a re-colonisation during that time are higher. Additionally, this region may have had established *R. exulans* populations before the Austronesian expansion. I expected the Normanby Island and New Guinea Polynesian rat populations to be particularly close genetically given their proximity, however Normanby Island is more similar to Sulawesi (F_{ST} 0.093) and Halmahera (F_{ST} 0.055) than is it to New Guinea (F_{ST} 0.201). This could be because New Guinea is a unique case in that it is particularly large and therefore have many rat subpopulations and increased variation, and the Polynesian rat may have been well established on New Guinea before the Austronesian expansion, therefore having a separate set of alleles than those of the rats the Lapita people carried out of islands closer to the mainland.

Although Kayangel is relatively nearby the southeast Asian and Australasian islands, such as Halmahera, which is approximately 1070km away (comparable to Halmahera and Borneo, which are approximately 1200km apart), the Kayangel Polynesian rat population is more genetically different than those islands. This suggests that Kayangel was either not part of the movement patterns (such as trade) that likely occurred within the southeast Asian and Australasian regions, or the Polynesian rat population there is particularly susceptible to inbreeding (Fig. 4). Both of these points could be related to the island's very small size ($1.7km^2$). In addition to this, Kayangel displayed several unusual results in the Hardy-Weinberg analysis (the specimens of which were removed). Hardy-Weinberg disequilibrium can be caused by founder effects and inbreeding, although it is also possible that methodical errors occurred to specimens in the Kayangel population data. One reason for Kayangel being more genetically isolated could be the direction of ocean currents. The Pacific-Indonesian Throughflow currents move from the east/northeast towards the south/southwest direction (Auer et al., 2019), meaning sailing towards Kayangel from the west or southwest would take more effort than sailing with the currents.

Central to this dataset appears to be Late Island (Fig. 1, Fig. 3, Fig. 6), which shares relatively close relationships with Wake Island to the north, the eastern Polynesian Islands, New Zealand to the south and Australasian islands back west. In some cases, the eastern islands are more genetically similar to the *R. exulans* population on Late Island than the rat populations on nearby islands, such as Mohotani – Late Island (F_{ST} 0.303, 3960.73km) compared to Tahanea – Mohotani (F_{ST} 0.565, 1002.45km). Based on this it could be suggested that the Austronesian expansion, following the movement through Melanesia, either radiated out from the Tongan region/Near Polynesia towards the north, east and south, or alternatively, trade operated more frequently with the islands more central to the Pacific region such as Late Island. Current literature supports that the movement expanded out from the Cook Island region further east (West et al., 2017; Horsburgh & McCoy, 2017; Ioannidis et al., 2021). In this dataset Late Island is the closest to this region, and more sampling from the Cook Island area in the future would help to support or dispute this hypothesis.

For island pairs that are more remote, such as Mohotani and Hatutaa, the close genetic relationships between them are likely a remnant of a more recent common ancestor (such as a shared parent who hopped islands), due to less time being isolated and being towards the end of the island movement trend. Another possibility is that trade between the local islands continued to occur frequently, yet long distance trade was much rarer, therefore gene flow was more likely to occur between the close remote islands only, although not in all cases as mentioned above. On top of this, the remote islands already had a smaller gene pool to begin with, due to the founder effect as described in the following paragraph. Particularly curious island relationships given their distance apart include mainland – Great Mercury Island (F_{ST} 0.343, fitted value 0.545, 9315.9km), Rakiura – Hatutaa (F_{ST} 0.313, fitted value 0.493, 6511.1km) and Rakiura – Mohotani (F_{ST} 0.307, fitted value 0.491, 6448.7km). The mainland – Great Mercury Island relationship could be explained by a more recent accidental transportation or perhaps by nucleotide substitutions "going backwards" and making the observed difference less than the true difference in variation. More likely however, is that the observed F_{ST} is inaccurate, given the *p*-value of 0.138 for this estimate. For Hatutaa and Mohotani, it is possible that ocean currents such as the South Pacific Gyre made the journey from Eastern Polynesia to New Zealand more achievable than would normally be expected given the distance.

It should be stated that it was not the aim of this project to find out every contributing factor to Polynesian rat colonisation and variation however, only if a correlation was present, therefore research into this is reserved for future studies.

How prevalent is increased homozygosity in Pacific R. exulans populations successively further from the mainland populations, and is inbreeding correlated with island size here?
There is a significant correlation between the level of inbreeding and the distance from the mainland. This supports successive founder events having shaped bottlenecks in the Polynesian rat populations and is in line with similar studies, such as (Ioannidis et al., 2021) who state on humans; "reconstruction of the branching Polynesian migration sequence reveals a serial founder expansion, characterized by directional loss of variants…".

However there is a lot of variance that is not explained by the distance from the mainland, nor by island area. Island area and distance from the mainland have low level collinearity (Fig. 4). Islands which are close to the mainland and/or are part of a continental shelf such as Borneo on the Sunda shelf, are generally larger, while isolated oceanic islands which were formed from volcanic activity for example, tend to be smaller. This distance-area collinearity means the two variables do not explain exactly the same thing but are very similar and this interferes with the GLM, as can be seen by the significant results in the single linear models but non-significant results in the combined GLM. Low R^2 indicates a low amount of the variance is explained, therefore there may be more dynamics that can help explain the rat genetics.

There are also island populations which fit the model very poorly, for example New Guinea and Normanby Island, which were both identified as outliers in the homozygosity linear models. As previously mentioned, they present an interesting case, because New Guinea is the biggest island in this dataset, and Normanby Island is immediately off the coast of New Guinea. They are represented in the data by only three (New Guinea) and one (Normanby Island) specimen, which combined have above 50% SNPs missing. New Guinea is also known to have many other rat species, such as the New guinea rat *R. novaeguineae*, and Van

Deusen's Rat *R. vandeuseni* (IUCN, 2019), which may also have an unseen effect on the rat diversity. Doubtful Sound in New Zealand is another potential outlier given the size of the South Island (Te Waipounamu), and the population sample only consisting of one specimen. Competition from other rats such as *R. rattus* and *R. norvegicus*, or other predators such as stoats could put pressure on the Polynesian rat and prevent it from diversifying. This was not the case when the Polynesian rat first came to New Zealand however, as the other introduced predators were brought by the Europeans. Many Polynesian rat populations on mainland New Zealand are already extinct due to competition from other species of rat and mouse (Taylor, 1975; Russell et al., 2015), and eradication campaigns (Towns & Broome, 2003; Russell & Broome, 2016).

Limitations

There is a variety of limitations to this project, in the areas of data, analysis and biology. The number of initially sampled specimens (478) is not in and of itself inadequate, however these were only sampled from 25 islands, which is too few to make a sound model with high power, for example when testing homozygosity per island. In particular I noticed that Late Island is the only island in the dataset from the Fijian and Tongan region. Areas that could have had greater sampling to fill in the gaps are Hawaii, New Caledonia, Samoa, Fiji and surrounding islands, the Cook Islands, Easter Island, Indonesia, for example Java, and more specimens from the mainland to have a better idea of the structure of the mainland population (which would also allow for a better STRUCTURE analysis of the source populations that gave rise to the current distribution of the Polynesian rat). Some of these 25 islands in the dataset had only a few specimens representing them, and this number may have been reduced during the data clean-up of the raw data, which means an island (such as Luzon in this dataset) may only be represented by one specimen. This is not ideal because that specimen may not be a good representative of the population. In addition, the populations of Polynesian rats may not be delimited by the island boundaries, for example New Guinea may have multiple subpopulations, or islands from the same atoll such as Tetiaroa may be fundamentally the same population despite living on different islands.

Sampling bias may have been present, for example islands which are easier to reach, accessible regions of the islands, multiple sampling from the same area and therefore perhaps the same family. Ideally measures should be taken to ensure the sampling is acceptably random, so that diversity is accounted for on every island. Island area (in kilometres squared) was modelled with distance from the mainland against the level of inbreeding, however their degree of collinearity means they essentially are different ways of explaining the same situation. The addition of more variables to help explain the increased of lack of variation between islands would help shed light on the situation. Distance from the mainland (for homozygosity testing and F_{ST}) I believe is not an ideal variable (especially on its own), because it does not truly describe how difficult it is to reach the island in all cases, for example wind and current patterns may make a certain island more or less difficult to reach. Interference from other factors may also play a role in the genetic variation or lack thereof on islands, for example the presence of eradication campaigns on the island, such as is common in New Zealand.

Conclusion

The aim of this research project was to characterise the genetic signatures of hitchhiking Polynesian rats along the Austronesian expansion by assessing the relationship between

pairwise genetic and geographic distances, and the level of inbreeding of *R. exulans* on islands of the Pacific successively further away from mainland southeast Asia, the primary methods of which were pairwise distance matrices and homozygosity assessment supported by linear modelling. It was expected that the closer the islands are to each other, the more similar the population genetics were, and the further from the mainland, the greater the level of inbreeding. These hypotheses are supported by the results of this research. This outcome supports the current literature and adds a new facet to the combined endeavours of the scientific communities of biological, archaeological and linguistic disciplines to unravel the complex nature of the movement of *R. exulans* and by proxy, humans, across the Pacific ocean during and following the Austronesian expansion. I recommend that in future studies, a broader range of islands are sampled from and additional explanatory variables are selected to help explain the reasons for genetic differences between islands.

References

Adler GH, Levins R. 1994. The Island Syndrome in Rodent Populations. *The Quarterly Review of Biology* 69:473–490. DOI: 10.1086/418744.

Athens JS. 2009. Rattus exulans and the catastrophic disappearance of Hawai'i's native lowland forest. *Biological Invasions* 11:1489–1501. DOI: 10.1007/s10530-008-9402-3.

Auer G, De Vleeschouwer D, Smith RA, Bogus K, Groeneveld J, Grunert P, Castañeda IS, Petrick B, Christensen B, Fulthorpe C, Gallagher SJ, Henderiks J. 2019. Timing and Pacing of Indonesian Throughflow Restriction and Its Connection to Late Pliocene Climate Shifts. *Paleoceanography and Paleoclimatology* 34:635–657. DOI: 10.1029/2018PA003512.

Barnes SS, Matisoo-Smith E, Hunt TL. 2006. Ancient DNA of the Pacific rat (Rattus exulans) from Rapa Nui (Easter Island). *Journal of Archaeological Science* 33:1536–1540. DOI: 10.1016/j.jas.2006.02.006.

Bryant D, Moulton V. 2004. Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution* 21:255–265. DOI: 10.1093/molbev/msh018.

Campbell DJ, Moller H, Ramsay GW, Watt JC. 1984. Observations on Foods of Kiore (Rattus exulans) found in Husking Stations on Northern Offshore Islands of New Zealand. *New Zealand Journal of Ecology* 7:131–138.

Carson MT, Hung H, Summerhayes G, Bellwood P. 2013. The Pottery Trail From Southeast Asia to Remote Oceania. *The Journal of Island and Coastal Archaeology* 8:17–36. DOI: 10.1080/15564894.2012.726941.

Collerson KD, Weisler MI. 2007. Stone Adze Compositions and the Extent of Ancient Polynesian Voyaging and Trade. *Science* 317:1907–1911. DOI: 10.1126/science.1147013.

Earl DA, VonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4:359–361. DOI: 10.1007/s12686-011-9548-7.

Edwards AWF. 2008. G. H. Hardy (1908) and Hardy–Weinberg Equilibrium. *Genetics* 179:1143–1150. DOI: 10.1534/genetics.104.92940.

Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* 14:2611–2620. DOI: 10.1111/j.1365-294X.2005.02553.x.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* 10:564–567. DOI: 10.1111/j.1755-0998.2010.02847.x.

Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package).

van der Geer AAE. 2018. Changing invaders: trends of gigantism in insular introduced rats. *Environmental Conservation* 45:203–211. DOI: 10.1017/S0376892918000085.

van der Geer AAE. 2019. Effect of isolation on coat colour polymorphism of Polynesian rats

in Island Southeast Asia and the Pacific. *PeerJ* 7:e6894. DOI: 10.7717/peerj.6894.

van der Geer AAE. 2020. Size matters: micro-evolution in Polynesian rats highlights body size changes as initial stage in evolution. *PeerJ* 8:e9076. DOI: 10.7717/peerj.9076.

Geiger M, Sánchez-Villagra MR, Lindholm AK. 2018. A longitudinal study of phenotypic changes in early domestication of house mice. *Royal Society Open Science* 5:172099. DOI: 10.1098/rsos.172099.

Gerlach G, Jueterbock A, Kraemer P, Deppermann J, Harmand P. 2010. Calculations of population differentiation based on GST and D: forget GST but not all of statistics! *Molecular Ecology* 19:3845–3852. DOI: 10.1111/j.1365-294X.2010.04784.x.

Gibbs GW. 2009. The end of an 80-million year experiment: a review of evidence describing the impact of introduced rodents on New Zealand's 'mammal-free' invertebrate fauna. *Biological Invasions* 11:1587–1593. DOI: 10.1007/s10530-008-9408-x.

Glaubitz JC. 2004. convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. *Molecular Ecology Notes* 4:309–310. DOI: 10.1111/j.1471-8286.2004.00597.x.

Hardy GH. 1908. Mendelian Proportions in a Mixed Population. *Science* 28:49–50. DOI: 10.1126/science.28.706.49.

Harper GA, Bunbury N. 2015. Invasive rats on tropical islands: Their population biology and impacts on native species. *Global Ecology and Conservation* 3:607–627. DOI: 10.1016/j.gecco.2015.02.010.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of molecular evolution* 22:160–74. DOI: 10.1007/BF02101694.

Horsburgh K, McCoy M. 2017. Dispersal, Isolation, and Interaction in the Islands of Polynesia: A Critical Review of Archaeological and Genetic Evidence. *Diversity* 9:1–21. DOI: 10.3390/d9030037.

Huson DH, Bryant D. 2006. Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23:254–267. DOI: 10.1093/molbev/msj030.

Ioannidis AG, Blanco-Portillo J, Sandoval K, Hagelberg E, Barberena-Jonas C, Hill AVS, Rodríguez-Rodríguez JE, Fox K, Robson K, Haoa-Cardinali S, Quinto-Cortés CD, Miquel-Poblete JF, Auckland K, Parks T, Sofro ASM, Ávila-Arcos MC, Sockell A, Homburger JR, Eng C, Huntsman S, Burchard EG, Gignoux CR, Verdugo RA, Moraga M, Bustamante CD, Mentzer AJ, Moreno-Estrada A. 2021. Paths and timings of the peopling of Polynesia inferred from genomic networks. *Nature* 597:522–526. DOI: 10.1038/s41586-021-03902-8.

IUCN. 2019.IUCN Red List of Threatened Species. Version 2019.3. *Available at www.iucnredlist.org* (accessed January 1, 2020).

Johnson AD. 2010. An extended IUPAC nomenclature code for polymorphic nucleic acids. *Bioinformatics* 26:1386–1389. DOI: 10.1093/bioinformatics/btq098.

Kirch P V. 2000. *On the road of the winds: an archaeological history of the Pacific islands before European contact*. Berkeley: University of California Press.

Larson G, Cucchi T, Fujita M, Matisoo-Smith E, Robins J, Anderson A, Rolett B, Spriggs M,

671 Dolman G, Kim T-H, Thuy NTD, Randi E, Doherty M, Due RA, Bollt R, Djubiantono
672 T, Griffin B, Intoh M, Keane E, Kirch P V., Li K-T, Morwood M, Pedriña LM, Piper PJ,
673 Rabett RJ, Shooter P, Van den Bergh G, West E, Wickler S, Yuan J, Cooper A, Dobney
674 K. 2007. Phylogeny and ancient DNA of Sus provides insights into neolithic expansion
675 in Island Southeast Asia and Oceania. *Proceedings of the National Academy of Sciences*
676 104:4834–4839. DOI: 10.1073/pnas.0607753104.

677 Lischer HEL, Excoffier L. 2012. PGDSpider: an automated data conversion tool for
678 connecting population genetics and genomics programs. *Bioinformatics* 28:298–299.
679 DOI: 10.1093/bioinformatics/btr642.

680 Lum JK, Cann RL. 1998. mtDNA and language support a common origin of Micronesians
681 and Polynesians in Island Southeast Asia. *American Journal of Physical Anthropology*
682 105:109–119. DOI: 10.1002/(SICI)1096-8644(199802)105:2<109::AID-
683 AJPA1>3.0.CO;2-Q.

684 MacArthur RH, Wilson EO. 1967. *The Theory of Island Biogeography*. Princeton, NJ, USA:
685 Princeton University Press. DOI: 10.1515/9781400881376.

686 Mantel N. 1967. The detection of disease clustering and a generalized regression approach.
687 *Cancer research* 27:209–220.

688 Matisoo-Smith E. 1994. The Human Colonisation of Polynesia. A Novel Approach: Genetic
689 Analyses of the Polynesian Rat (Rattus exulans). *The Journal of the Polynesian Society*
690 103:75–87.

691 Matisoo-Smith E. 2015. Tracking Austronesian expansion into the Pacific via the paper
692 mulberry plant. *Proceedings of the National Academy of Sciences* 112:13432–13433.
693 DOI: 10.1073/pnas.1518576112.

694 Matisoo-Smith E, Roberts RM, Irwin GJ, Allen JS, Penny D, Lambert DM. 1998. Patterns of
695 prehistoric human mobility in Polynesia indicated by mtDNA from the Pacific rat.
696 *Proceedings of the National Academy of Sciences* 95:15145–15150. DOI:
697 10.1073/pnas.95.25.15145.

698 Matisoo-Smith E, Robins J. 2004. Origins and dispersals of Pacific peoples: Evidence from
699 mtDNA phylogenies of the Pacific rat. *Proceedings of the National Academy of Sciences*
700 101:9167–9172. DOI: 10.1073/pnas.0403120101.

701 Matisoo-Smith E, Robins J. 2009. Mitochondrial DNA evidence for the spread of Pacific rats
702 through Oceania. *Biological Invasions* 11:1521–1527. DOI: 10.1007/s10530-008-9404-
703 1.

704 Mirabal S, Cadenas AM, Garcia-Bertrand R, Herrera RJ. 2013. Ascertaining the role of
705 Taiwan as a source for the Austronesian expansion. *American Journal of Physical*
706 *Anthropology* 150:551–564. DOI: 10.1002/ajpa.22226.

707 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of Population Structure Using
708 Multilocus Genotype Data. *Genetics* 155:945–959. DOI: 10.1093/genetics/155.2.945.

709 Roberts M. 1991. Origin, dispersal routes, and geographic distribution of Rattus exulans, with
710 special reference to New Zealand. *Pacific Science* 45:123–130.

711 Roullier C, Benoit L, McKey DB, Lebot V. 2013. Historical collections reveal patterns of
712 diffusion of sweet potato in Oceania obscured by modern plant movements and
713 recombination. *Proceedings of the National Academy of Sciences* 110:2205–2210. DOI:

10.1073/pnas.1211049110.

Russell JC., Broome KG. 2016. Fifty years of rodent eradications in New Zealand. *New Zealand Journal of Ecology* 40:197–204.

Russell JC, Caut S, Anderson SH, Lee M. 2015. Invasive rat interactions and over-invasion on a coral atoll. *Biological Conservation* 185:59–65. DOI: 10.1016/j.biocon.2014.10.001.

Semagn K, Babu R, Hearne S, Olsen M. 2014. Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Molecular Breeding* 33:1–14. DOI: 10.1007/s11032-013-9917-x.

Taylor RH. 1975. What limits kiore (Rattus exulans) distribution in New Zealand? *New Zealand Journal of Zoology* 2:473–477. DOI: 10.1080/03014223.1975.9517888.

Thomson V, Aplin KP, Cooper A, Hisheh S, Suzuki H, Maryanto I, Yap G, Donnellan SC. 2014. Molecular Genetic Evidence for the Place of Origin of the Pacific Rat, Rattus exulans. *PLoS ONE* 9:e91356. DOI: 10.1371/journal.pone.0091356.

Towns DR, Broome KG. 2003. From small Maria to massive Campbell: Forty years of rat eradications from New Zealand islands. *New Zealand Journal of Zoology* 30:377–398. DOI: 10.1080/03014223.2003.9518348.

Weinberg W. 1908. Über den Nachweis der Vererbung beim Menschen. *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg* 64:369–382.

Weisler MI, Bolhar R, Ma J, St Pierre E, Sheppard P, Walter RK, Feng Y, Zhao J, Kirch P V. 2016. Cook Island artifact geochemistry demonstrates spatial and temporal extent of pre-European interarchipelago voyaging in East Polynesia. *Proceedings of the National Academy of Sciences* 113:8150–8155. DOI: 10.1073/pnas.1608130113.

West K, Collins C, Kardailsky O, Kahn J, Hunt TL, Burley D V., Matisoo-Smith E. 2017. The Pacific Rat Race to Easter Island: Tracking the Prehistoric Dispersal of Rattus exulans Using Ancient Mitochondrial Genomes. *Frontiers in Ecology and Evolution* 5:1–13. DOI: 10.3389/fevo.2017.00052.

Wigginton JE, Cutler DJ, Abecasis GR. 2005. A Note on Exact Tests of Hardy-Weinberg Equilibrium. *The American Journal of Human Genetics* 76:887–893. DOI: 10.1086/429864.

Wilkins AS, Wrangham RW, Fitch WT. 2014. The “Domestication Syndrome” in Mammals: A Unified Explanation Based on Neural Crest Cell Behavior and Genetics. *Genetics* 197:795–808. DOI: 10.1534/genetics.114.165423.

Wilmshurst JM, Anderson A, Higham TFG, Worthy TH. 2008. Dating the late prehistoric dispersal of Polynesians to New Zealand using the commensal Pacific rat. *Proceedings of the National Academy of Sciences* 105:7676–7680. DOI: 10.1073/pnas.0801507105.

Wilmshurst JM, Hunt TL, Lipo CP, Anderson A. 2011. High-precision radiocarbon dating shows recent and rapid initial human colonization of East Polynesia. *Proceedings of the National Academy of Sciences* 108:1815–1820. DOI: 10.1073/pnas.1015876108.

Wright S. 1978. *Evolution and the Genetics of Populations, Variability Within and Among Natural Populations*. Chicago: University of Chicago Press.