Title: How Good are Indirect Tests at Detecting Recombination in Human mtDNA?

Authors: Daniel James White *, David James Bryant **§, Neil John Gemmell ***§

Current affiliations:
* Biodiversity and Informatics, Landcare Research New Zealand, Auckland, New Zealand 1072
** Department of Mathematics and Statistics, University of Otago, Dunedin, New Zealand 9054
*** Centre for Reproduction and Genomics, Department of Anatomy, University of Otago, Dunedin, New Zealand 9054
§ Allan Wilson Centre for Molecular Ecology and Evolution, University of Otago, Dunedin, New Zealand 9054
Running head: Can we detect human mtDNA recombination?

Key words: Recombination, mtDNA, Indirect Tests, Effectiveness, Human

Corresponding author: Daniel James White
Landcare Research New Zealand
PO Box 92170
Auckland Mail Centre 1142
231 Morrin Road, St. John’s
Auckland, New Zealand
phone: +64 9 574 4161
fax: +64 9 574 4101
email: whited@landcareresearch.co.nz
Abstract

Empirical proof of human mtDNA recombination in somatic tissues was obtained in 2004; however, there remains a lack of irrefutable evidence for recombination in human mtDNA at the population level. Our inability to demonstrate convincingly a signal of recombination in population data sets of human mtDNA sequence may be due, in part, to the ineffectiveness of current indirect tests. Previously, we tested some well-established indirect tests of recombination (LD versus distance using $D'$ and $r^2$, Homoplasy Test, PHI, NSS and Max Chi Squared) on sequence data derived from the only empirically confirmed case of human mtDNA recombination thus far, and demonstrated that some methods were unable to detect recombination. Here, we assess the performance of these six well established tests and explore what characteristics specific to human mtDNA sequence may affect their efficacy, by simulating sequence under various parameters with levels of recombination ($\rho$) that vary around an empirically derived estimate for human mtDNA (population parameter $\rho = 5.492$). No test performed infallibly under any of our scenarios, and error rates varied across tests, whereas detection rates increased substantially with $\rho$ values > 5.492. Under a model of evolution that incorporates parameters specific to human mtDNA, including rate heterogeneity, population expansion, and $\rho = 5.492$, successful detection rates are limited to a range of 7 to 70% across tests with an acceptable level of false positives: the NSS incompatibility test performed best overall under these parameters. Population growth seems to have the greatest impact on recombination detection probabilities across all models tested, likely due to its impact on sequence diversity. The implications of our findings on our current understanding of mtDNA recombination in humans are discussed.
Introduction

If recombination is occurring in human mitochondrial DNA (mtDNA) at a sufficiently high level then many evolutionary analyses based on the molecule (Anisimova et al. 2003; Arenas and Posada 2010a; b; Schierup and Hein 2000a; b), and indeed the models of evolution of the molecule itself, will need to undergo substantial revision. The question of whether recombination occurs at this level is, however, largely disregarded by the scientific community. This consensus has been reached based on a number of factors. It was long held that the mitochondrial DNA molecule evolves in a strictly clonal manner, with exclusively uniparental inheritance and no recombination. In fact, originally, it was not clear whether enzymes required for recombination were present in mitochondria at all. Further, it was accepted that there did not exist significant enough mitochondrial heteroplasmy (more than one mitotype present in a cell or individual) for there to be any biological impact, even if recombination were to occur. Early studies that reported apparent signals of recombination in human mtDNA (Awadalla et al. 1999; Eyre-Walker et al. 1999; Hagelberg et al. 1999) were largely discounted based on statistical arguments (Innan and Nordborg 2002; Jorde and Bamshad 2000; Kumar et al. 2000) and inaccuracies in sequence data (Hagelberg et al. 2000; Kivisild and Vilems 2000).

However, the body of evidence supporting the potential for biologically important recombination in human mtDNA has mounted over recent years. For example, in 1999 it was discovered that the enzymatic machinery necessary for recombination is present in mitochondria (Lakshmipathy and Campbell 1999; Thyagarajan et al. 1996). Important revelations came in 2004 and 2005, when widespread mtDNA recombination was discovered in up to 14.2% and 16.1% of
animal populations, respectfully (Piganeau et al. 2004; Tsaousis et al. 2005), although a recent and interesting study has shown that positive results for recombination in animal mtDNA could be an artefact of a specific form of mutation rate heterogeneity (Sun et al. 2011). Experimental work has shown that recombination can occur between heterologous mitochondrial molecules, harbouring pathogenic mutations, in human cytoplasmic hybrid cells (D'aurelio et al. 2004). Finally, studies on the patterns of tetraplasmy in doubly heteroplasmic individuals not only suggest mitochondrial recombination to be relatively common in human skeletal muscle tissue, but also demonstrate the inheritability of recombinant mtDNA molecules (Zsurka et al. 2007; Zsurka et al. 2005).

For recombination to have a biological impact it must occur between two different mitotypes, and thus a state of mitochondrial heteroplasmy is required. To this end, it is now accepted that mitochondrial heteroplasmy is a regular feature among higher eukaryotes (White et al. 2008). As evidence, the paternal inheritance of mtDNA, a common source of heteroplasmcy, is relatively widespread across the animal kingdom (Fontaine et al. 2007; Gyllensten et al. 1991; Sherengul et al. 2006; Zhao et al. 2004), and is also known to occur in humans (Schwartz and Vissing 2002; Zsurka et al. 2005). Indeed, human mitochondrial heteroplasmy can persist in the germ line over several generations (Ivanov et al. 1996).

The most convincing direct evidence of a recombination event in human mtDNA to date remains that of a male sufferer of a mitochondrial myopathy, who had a 10:1 ratio of paternal to maternal mtDNA in his skeletal muscle. A deletion in his paternally derived mtDNA gave rise to the disorder, and it was found that 0.7% of mtDNA molecules were recombinant (Kraytsberg et al. 2004). This research group
subsequently went on to show that recombination in this tissue might be relatively common (Zsurka et al. 2005).

To understand the appropriate role of recombination in evolutionary analyses centred around the mtDNA molecule, we require more than sporadic cases leading to dysfunction. Instead, we require evidence for mitochondrial recombination at a population level. Such evidence can be provided by the numerous statistical tests that estimate the probability that polymorphism distributions in DNA sequence are due to recombination (Bruen et al. 2006; Posada and Crandall 2001, see http://www.bioinf.manchester.ac.uk/recombination/programs.shtml for an up to date list). The early evidence for mitochondrial recombination in human populations using these methods (Awadalla et al. 1999; Eyre-Walker et al. 1999) is considered equivocal at best, however, and has largely been disregarded. The reasons for refuting the positive signals for recombination as revealed in these studies include the incorrect choice of statistical test (Jorde and Bamshad 2000; Kumar et al. 2000), the quality of sequence data (Kivisild and Villems 2000), and inaccurate interpretation of results (Innan and Nordborg 2002). Subsequent attempts to repeat the findings of the first studies have been unsuccessful (Elson et al. 2001; Piganeau and Eyre-Walker 2004).

The question we focus on here is why there is such a lack of compelling evidence for mtDNA recombination frequency in human populations, given that there is definitely potential for biologically relevant mitochondrial recombination to occur, and it is occurring at the population level in our animal counterparts. With a lack of compelling evidence, the majority of the scientific community is content to accept that it may simply just not be occurring at a significantly high enough level for due concern.
One intriguing possibility is that recombination may well be occurring in human mtDNA, but that we are not actually able to detect it with the current set of tools. In a previous study we demonstrated that some of the most robust indirect statistical tests for recombination detection are not effective at detecting recombination in human mtDNA, even when it is known to have occurred (White and Gemmell 2009).

Here, we explore how some of the processes under which human mtDNA evolves may affect the effectiveness of indirect tests of recombination detection. Specifically, we evolve DNA sequence under parameters consistent with human mtDNA, including rate heterogeneity and population expansion. Simulations are performed using variable levels of recombination to determine the limitations of six well established indirect tests of recombination. The power of these recombination tests has been comprehensively assessed elsewhere (Bruen et al. 2006; Posada and Crandall 2001; Wiuf et al. 2001) and our objective here is not to repeat such work. Rather, our intent is to assess the effects, on the detection abilities of popular tests of recombination, of some of the key parameters known to impact the evolution of this molecule in humans. In particular, we ask the question, what level of recombination in human mtDNA might we need before it can be reliably detected with the currently available recombination detection tools? We find that current indirect tests of recombination perform very poorly even under our simple models of sequence evolution, and likely will perform even worse as model complexity increases towards a scenario closer to biological reality. We discuss how our findings could influence previous studies, and current opinions, of human mtDNA recombination.
Materials and Methods

Sequence Simulation

Sequence was simulated under a relatively simple scenario consistent with the evolution of human mtDNA, with varying levels of recombination. To estimate the population parameters to use for the simulation of sequences, recombinant mtDNA sequences from the only known case of empirical recombination detection in human mtDNA (Kraytsberg et al. 2004) were run in LAMARC v2.1 (Kuhner 2006). Assuming a mutation rate, $\mu$, of $5 \times 10^{-7}$ per site per generation for the mtDNA molecule (Mishmar et al. 2003), the required number of individuals to generate the observed genetic diversity within this sample of recombined sequences, effective population size $N$, was approximated at 1370 (from $N = \theta/\mu$, where $\theta$ is the mutation rate parameter). The recombination factor, $R$, was estimated in LAMARC from $R = r/\mu$, where $r$ is the recombination rate per inter-site link per generation and $\mu$ is the mutation rate per site per generation (Kuhner 2006), and can be described as the relative frequency of recombination next to a given site compared to mutation at that site. For our sequences, and assuming $\mu$ to be $5 \times 10^{-7}$, $R$ was estimated at 1.170 giving an $r$ value of $5.8485 \times 10^{-7}$ (see White and Gemmell 2009 for more details).

These parameters were incorporated into RECODON v1.5.0, which generates sequence under a coalescent model (Arenas and Posada 2007). Population parameters per locus were estimated from a product of twice the number of inheritable copies, the per-site factor and the length in basepairs of the locus, $l$. As such, the per-locus population recombination parameter, $\rho$, was approximately $Nrl$ for mitochondrial sequence when we assume an equal number of females and males in the population.
(Arenas and Posada 2007); we set l to 6854bp, which gives a value for ρ of 5.492 for the empirical data set. To account for recombination occurring only in females for mtDNA, the per-site recombination rate was halved for the simulations, otherwise the level of recombination is overestimated in the simulated sequences. Three models of sequence evolution were used: (i) constant population size and homogeneous mutation rate; (ii) constant population size and heterogeneous mutation rate; and (iii) population growth and heterogeneous mutation rate. For models that included rate heterogeneity a gamma distribution of mutation rates with alpha set at 0.2 was used (Yang 1996), and for the model that included population growth an exponential growth rate of $1 \times 10^{-3}$ per generation was incorporated (Pluzhnikov et al. 2002). For each evolutionary model, ρ was set to 0, 2, 5.492, 8, 10, and 12 to span the estimated value for mitochondrial DNA (5.492), and 100 samples of 33 sequences were generated for each level of recombination. Sequences were evolved under the Jukes-Cantor nucleotide model assuming equal frequencies for all nucleotides, with an equal transition-transversion ratio.

The Tests

Six well-known indirect tests of recombination were chosen for assessment. These include three population-based tests: the correlation between linkage disequilibrium (LD) and physical distance using the D’ measure of LD; the correlation between LD and physical distance using the $r^2$ measure of LD; and the Homoplasy Test (Smith and Smith 1998). Three general methods were also assessed, including Max Chi Squared (Smith 1992), the NSS (Jakobsen and Easteal 1996), and the more recently developed PHI test (Bruen et al. 2006). We have described these
indirect tests of recombination in some detail elsewhere (White and Gemmell 2009),
but briefly re-encapsulate the key points of difference among them here.

The use of LD to predict recombination relies on a significant negative
correlation between physical distance and strength of LD in the presence of
recombination (Awadalla et al. 1999). For both measures of LD, Pearson’s correlation
coefficient was used, and the statistical significance of the correlation was estimated
after 1000 random permutations of the data using a Mantel test, all implemented in
RecombiTEST (Piganeau et al. 2004). The Homoplasy Test looks at whether there are
more homoplasies on a phylogenetic tree than expected under a model of clonal
inheritance, where a homoplasy is the co-occurrence of a polymorphism on separate
branches of the tree (Smith and Smith 1998). We implemented this test in the Linux
operating system using a C translation of the original QBasic version, an unsupported
version kindly provided by David Posada (University of Vigo). To simulate the
process of synonymous site selection, which is a recommended step to control for the
compounding effects of selection on recombination detection, a second file was
generated for analysis, using every third base pair of the 6854 bp-long simulated
sequence. The C version of Homoplasy was validated by comparing its performance
to the original QBasic program using 5 files with no recombination (ρ=0), and 5 files
with an extreme level of recombination (ρ=15). Results are shown in supporting table
S1. The Max Chi Squared test compares the arrangement of segregating sites either
side of a putative crossover break point. Both the NSS and PHI tests look at the
distribution of incompatible sites, where sites are compatible only if their history
includes no recurrent or convergent mutation, or recombination. C versions of Max
Chi Squared, NSS and PHI (http://www.maths.otago.ac.nz/~dbryant/software.html)
were run in Linux, and the significance of results estimated by randomly positioning
the informative sites 1000 times and determining the proportion of permuted scores that are: (i) below the observed test score, (ii) at least as high as the observed test score, and (iii) less than or equal to the observed score, respectfully. For all six tests a significant signal for recombination was assumed for a dataset if a p value of 0.05 or less was achieved. For more detail on the tests see White and Gemmell (2009), and supporting file S1.

**Results**

For the Kraytsberg recombinant sequence data, recombination per site per generation was estimated by LAMARC as $5.849 \times 10^{-7}$, which gives a most probable population recombination parameter per gene, $\rho$, of 5.492. The population mutation parameter, $\theta (N\mu)$, was estimated as $6.850 \times 10^{-4}$, which equates to 4.695 per gene.

**Effect of Recombination Rate**

For sequence simulated under the simplest scenario of no population growth or mutation rate heterogeneity, recombination detection was at, or near, 100% for $\rho=12$ for the majority of tests (see figure 1a). The clear exception to this is the Max Chi Squared test, which reported recombination in only 87% of cases. For the lowest level of recombination with $\rho$ set to 2, only one test (the Homoplasy Test) detected recombination more than 90% of the time. Overall, levels of detection were relatively low for $\rho = 2$, ranging from 58 to 99%, with a mean of 74.0% (compared with a mean of 98.0% for $\rho = 12$). Interestingly, for $\rho = 5.492$ (the value estimated for the
empirical example of human mtDNA recombination) the rates of detection ranged from 81 to 100% across all tests, with a mean of 94.4% (figure 1b).

Although there was a slight reduction in efficiency of recombination detection when only 3rd positions were considered in the Homoplasy Test, compared with when all sites are used, this difference was not significant.

Five of the six tests reported some level of recombination when there was none. Only the PHI test did not. The percentage of false positives ranged from 1% for NSS to 30% with the Homoplasy Test using all sites.

Effect of Mutation Rate Heterogeneity

When mutation rate heterogeneity, described by a gamma distribution with α equal to 0.2, is incorporated into sequence evolution, a very similar pattern to the simple scenario is observed in rates of detection across all tests of recombination. When means of detection rates across all tests are compared between the two scenarios, there is no significant difference (p < 0.70, paired student’s t-test). When results are compared between the two scenarios for each test separately, Max Chi Squared showed a significant drop in detection rates at all levels of recombination with mutation rate heterogeneity (p < 0.009, paired student’s t-test, table 1). LD versus distance using r² also showed a smaller, yet significant, reduction in detection rates with mutation rate heterogeneity (p < 0.042, paired student’s t-test, table 1).

Effect of Mutation Rate Heterogeneity and Population Growth

When a population growth rate of 1 x 10⁻³, and mutation rate heterogeneity described by a gamma distribution with α equal to 0.2, were incorporated into simulations, detection rates were significantly lower (p < 8.18 x 10⁻³, paired student’s
t-test on means for all tests at each value of $\rho$, see figure 2). All tests were less able to detect recombination at all levels of recombination. Although the Homoplasy Test maintained the highest detection rate, even at low levels of recombination (93% for all sites at $\rho = 2$), the rate of false positives was 61% for this test when all data were examined, and 37% when 3rd positions were used. At the other end of the spectrum, Max Chi Squared performed dismally with detection rates ranging from 4 to 11%.

The detection rates of the four remaining tests were similar, ranging from 50% (NSS) to 74% (LD versus distance using $D'$) at $\rho = 12$. At $\rho = 5.492$, the value estimated for the Kraytsberg recombinant sequence, rates of detection for these four tests ranged from 37% with LD versus distance using $r^2$, to 70% with LD versus distance using $D'$.

A possible reason for the difference in results between simulations without rate heterogeneity and population growth, compared with those that did include these parameters is the underlying sequence diversity. This has previously been shown to have a strong effect on the efficiency of indirect tests of recombination detection (Bruen et al. 2006; Posada and Crandall 2001; Wiuf et al. 2001). We calculated the mean nucleotide diversity, $\pi$, among sequences (i) in the simple scenario and (ii) with mutation rate heterogeneity and population growth as $2.392 \times 10^{-3}$ and $1.257 \times 10^{-3}$, respectively. The latter is comparable with the estimate of $\pi$ for the empirical data set, $1.250 \times 10^{-3}$ (White and Gemmell 2009). For simulated sequence with mutation rate heterogeneity and no population growth, $\pi$ among sequences was estimated as $2.545 \times 10^{-3}$.

To assess the impact of the model of sequence evolution on false positive error rates of the tests, histograms of p-values obtained for $\rho = 0$ were plotted for both the
most simple model of sequence evolution, and the model including population growth and rate heterogeneity were made (figure 2a and 2b, respectively). A uniform distribution of p-values indicates an unbiased test, a strongly right-skewed distribution indicates a conservative test, while a strongly left-skewed distribution indicates high false positive error rates. We observed great variability across tests, with the incompatibility tests (PHI and NSS) having distributions strongly skewed to the right whereas the Homoplasy test shows a left skewed distribution.

Discussion

In this paper we have explored the possibility that the lack of evidence for recombination in human mtDNA is not necessarily due to its absence, but rather might be a consequence of a lack of power to detect it among current tests. To test this thesis we simulated three sets of sequence with varying levels of recombination: (i) sequence generated under a simple model using a reliable estimate for the rate of mtDNA mutation (Mishmar et al. 2003), but no population growth or mutation rate heterogeneity; (ii) sequence generated using a model that incorporates an unequal distribution of mutation rates across human mtDNA molecules (Yang 1996); and (iii) sequence generated using a model that incorporates a non-uniformly distributed mutation rate, as well as rates of recent human population growth (Pluzhnikov et al. 2002).

Performance of Tests

Under the simple model of sequence evolution, the majority of tests were able to detect recombination at least 95% of the time when levels of recombination are
relatively high, i.e. with a population recombination parameter, \( \rho \), of at least 8. The exception to this includes the Max Chi Squared test for which a maximum rate of successful detection was estimated at 87%. When \( \rho \) is equal to 5.492, the estimated value for human mtDNA based on the Kraytsberg recombinant sequence data (White and Gemmell 2009), detection rates were lower for all tests except the Homoplasy Test, and fell as low as 81% for Max Chi Squared.

The performance of all tests changed dramatically when population growth and mutation rate heterogeneity were incorporated into the model of sequence evolution, but rate heterogeneity on its own had little effect (table 1). With the inclusion of an exponential growth factor into sequence simulation, there was a reduction in performance for all tests; most tests showed a marked reduction in their power of detection across all levels of recombination. This effect is most likely explained by the reduction in sequence diversity (e.g., from \( \pi = 2.39 \times 10^{-3} \) (s.d. 0.60 x \( 10^{-3} \)) in the simple scenario to \( 1.26 \times 10^{-3} \) (s.d. 0.41 x \( 10^{-3} \))), as the negative impact of low sequence diversity on the tests in this study has been demonstrated before (Bruen et al. 2006; Posada and Crandall 2001; Wiuf et al. 2001). Most severely affected was the Max Chi Squared test, which is unsurprising given its inefficiency in scenarios of low sequence diversity, particularly under population growth (Bruen et al. 2006; Posada and Crandall 2001; Wiuf et al. 2001). Thus, this test (using the implementation of Bruen et al. (2006)) can be excluded as a useful test for detecting recombination under these conditions.

An apparent exception to the overall trend of reduced power under population growth and mutation rate heterogeneity is the Homoplasy Test, which maintained levels of recombination detection above 90%, even for a \( \rho \) value of 2. The rate of false positives more than doubled for this test, however, to 61%. When only 3rd positions
are used, a step recommended in real data to prevent the compounding effects of selection on recombination detection (Piganeau and Eyre-Walker 2004; Smith and Smith 1998), the rate of false positive detection falls to 37% with the Homoplasy test, while simultaneously retaining relatively high power.

Of course, a test might have high power not because it is accurate, but because it rejects the null model unnecessarily. We can check for this by simulating with no recombination, and then checking which tests falsely detect recombination. Figure 2 clearly shows that the C version of the Homoplasy Test used here has an unacceptable false positive rate with our data set, meaning that the apparent low false negative rate (i.e. high power) of this test is artefactual. In contrast, the incompatibility tests are overly conservative for our sequences, which have similar nucleotide diversity levels to empirical mtDNA data, suggesting inflated false negative rates with a nominal p value cut-off of 0.05 for these tests.

When these results are compared with previous studies (Brown et al. 2001; Bruen et al. 2006; Posada and Crandall 2001; Wiuf et al. 2001) several comparisons can be made. Overall, the apparent power of tests to detect recombination, in a scenario of no population growth, is higher in this analysis than in previous studies (Bruen et al. 2006; Posada and Crandall 2001; Wiuf et al. 2001). For most tests, recombination is detected at higher percentages across all levels of recombination used here. This is in spite of the lower sequence divergence of our simulated sequences. It has previously been shown that the power of these tests increases with increasing sequence diversity (Bruen et al. 2006; Posada and Crandall 2001; Wiuf et al. 2001); however, we estimated an average pairwise sequence diversity, π, for the simple scenario sequences of $2.392 \times 10^{-3}$ (s.d. $0.60 \times 10^{-3}$); approximately five-fold
less than the least divergent sequences used in previous simulation studies (Bruen et al. 2006; Posada and Crandall 2001; Wiuf et al. 2001). Reasons for the higher rate of detection here, compared with previous studies, may therefore include a relatively large sample size in this study (n = 33), and a long sequence length (l = 6854 bp), both of which can increase the chance of recombination detection. However, the benefits of a long sequence length may be restricted by the incorporation of a sliding window into sequence analysis (Bruen et al. 2006; Wiuf et al. 2001).

In their comprehensive review, Posada and Crandall (2001) assessed 14 tests of recombination over a range of parameters, and demonstrated that the Homoplasy Test was the most powerful at lower levels of divergence (π ~ 1%). As the level of diversity in our simulated sequences is lower again, and more biologically relevant for studies in humans, an observation of relatively high power with the Homoplasy Test for our data would not be completely unexpected. Posada and Crandall (2001) also showed that, with low mutation rate heterogeneity and low sequence divergence, false positives should be low with the Homoplasy Test. However, we report a higher false positive rate than previous studies at 25%, which is substantially larger than expectation (at p = 0.05), for the scenario that included no mutation rate heterogeneity. This observation suggests a discrepancy between the C version of the Homoplasy test used in the Posada and Crandall’s 2001 study and that used here. To investigate this, we repeated the analyses done in the Posada and Crandall study, which used sequence simulated with lower diversity and shorter length, using our implemented C version. A similar false positive rate was found as in Posada and Crandall’s study (4% - data not shown). The inflated probability of false positives here is therefore due to other factors, such as our larger sequence length and larger effective population size. Larger sequence length could give rise to increased false
positives by the method in which the test accounts for homoplasies. With larger sequence length, and corresponding larger number of effective sites, the Homoplasy Test will consider the likelihood that any homoplasies present to have arisen by recurrent mutation to be very small, and instead will report recombination. The type 1 error rate rose to 61% when population growth and rate heterogeneity are included across all sites, and this rise is likely driven, at least in part, by the unequal distribution of mutation rates (Posada and Crandall 2001). The apparent high power with the Homoplasy Test reported here should therefore be treated with caution.

The PHI statistic of Bruen et al. (2006) has been reported as a powerful and robust test that can reliably detect recombination under a wide spectrum of parameters, including population growth. Moreover, it has been shown to report few false positives, even under strong mutation rate heterogeneity and autocorrelation of substitution rates. However, when there is little sequence diversity, it may be conservative in detecting recombination (Bruen et al. 2006). Our results agree with these initial findings in that the power of this test was good at high levels of recombination; however, its performance was reduced when recombination parameters were low. This reduced performance with low recombination is likely due to both reduced sequence diversity, and the incorporation of a sliding window that limits any benefits derived from the long sequence length (6854 bp). NSS, like PHI, is based on incompatibility, and incompatibility methods have been shown to be among the most powerful across a range of parameters (Brown et al. 2001; Posada and Crandall 2001; Wiuf et al. 2001). Here, it also appears to be successful for mtDNA, under no population growth and a uniform mutation rate. Our results agree with the results of earlier studies in that the incompatibility-based tests (PHI and NSS) are
conservative when sequence diversity is low, especially at low levels of recombination (Bruen et al. 2006; Posada and Crandall 2001).

The Max Chi Squared method was originally developed for detecting recombination between distinct DNA sequences, showing substantial divergence (Smith 1992). Under such a scenario it has been shown to be a powerful test (Posada and Crandall 2001). Due to the low diversity of our simulated sequences, its poor performance in this study is unsurprising. However, the implementation of the Max Chi Squared test we used employs the entire sequence and all possible Chi squares. If Chi square tables with expected values of two or less are excluded, a step that prevents overtly large Chi squares being produced when expected values are low, the Max Chi Squared test may become more powerful without inflation of type 1 error (Piganeau and Eyre-Walker 2004).

Historically, there has been some debate as to which measure of LD, $r^2$ or D’, is preferable for the detection of recombination. Indeed, some authors have questioned the use of LD decay as a reliable indication of recombination at all (Innan and Nordborg 2002). However, the LD association test performed favourably compared with four other tests under no population growth, in a more recent simulation study, using both $r^2$ and D’ (Bruen et al. 2006). While the first study to show recombination in human mtDNA using LD employed the $r^2$ measure (Awadalla et al. 1999), it has been argued that only D’, and not $r^2$, is a suitable measure of LD for mtDNA based on certain statistical criteria, such as for maximum values of $r^2$ to be achieved all alleles must have equal frequency (Jorde and Bamshad 2000; Kumar et al. 2000). In contrast, it has been suggested that the $r^2$ measure of LD is preferable over D’ for synonymous mutations in human mtDNA, as synonymous sites in human mtDNA are described well by the finite sites model (Piganeau and Eyre-Walker
2004), a model for which $r^2$ is more suited (Mcvean et al. 2002). Simulation studies have generally not been able to differentiate the relative efficacies of using D’ or $r^2$ (Bruen et al. 2006; Meunier and Eyre-Walker 2001). Like (Bruen et al. 2006) we found that $r^2$ is slightly less powerful under population growth, though another study (White and Gemmell 2009) gives an example where a test based on $r^2$ detected recombination in human mtDNA, when D’ did not; the only case in which we know recombination has occurred. It has been suggested that hypervariable mutation leading to recurrent mutation would cause a reduction in detection using D’, but not $r^2$ (Meunier and Eyre-Walker 2001), and as this parameter is not included in our simulations here (whereas other evolutionary factors are), hypervariability could explain the discrepancy between the results presented here and those of our earlier analyses of empirical and simulated data. It would be interesting to explore further the effect of recurrent mutation on these tests.

It is perhaps the scenario of mutation rate heterogeneity and population growth that is most relevant for human mtDNA, as these phenomena are believed to be major factors of human mtDNA evolution and recent demographic history (Pluzhnikov et al. 2002; Yang 1996). As only minor differences are seen when mutation rate heterogeneity alone is included in models of sequence evolution, it is probable that population growth, and the subsequent reduction in sequence diversity, is the major contributor to the marked reduction in efficiency for tests under a slightly more complex scenario. Thus, it appears that the parameters under which human mtDNA is evolving have a substantial impact on the performance of these indirect tests of recombination.
Ignoring the homoplasy test due to its inflated false positive rate, and considering the discrepancies in the LD D’ test with previous studies (White and Gemmell 2009), it is the incompatibility tests that are most robust to the parameters of sequence evolution implemented here, while maintaining low false positive rates. Since the PHI test was unable to detect recombination in empirical data (White and Gemmell 2009), the NSS test may be the most suited of the tests included here to detect recombination in human mtDNA. The LD r^2 test may also be useful.

The focus of this study, however, was not to provide estimates of the power of these tests to detect recombination, but rather to test how the parameters under which human mtDNA evolve might affect the efficacy of indirect tests of recombination. In general, the six approaches tested, all of which have been widely applied for the detection of recombination in sequence data, perform poorly. We demonstrate that under a realistic rate of recombination for human mtDNA, the performance of most indirect tests is reduced at least two-fold when population growth and mutation rate heterogeneity are considered compared with a scenario when these factors are absent. Further, it is apparent that error rates vary dramatically across tests and that relying on a universal nominal p value of 0.05 to demonstrate statistical significance is simply not appropriate. As the evolution of human mtDNA is likely to be more complex than we simulated here, we strongly recommend further development of indirect tests of recombination for accurate detection in human mtDNA. At the very least, we suggest that simulations be run to explore the limitations of tests as part of empirical investigations.

If we were to re-investigate past efforts to detect recombination in human populations, considering the results of this study, conclusions may need to be
reviewed. Studies thus far have employed several indirect tests, including those that we have shown to be unsuitable for human mtDNA, and evidence has generally been negative or inconclusive (Elson et al. 2001; Herrnstadt et al. 2002; Ingman et al. 2000; Jorde and Bamshad 2000; Piganeau and Eyre-Walker 2004). By removing non-suitable tests from studies, the landscape of evidence for no recombination in human populations changes. For example, in the most comprehensive screen for recombination in human mtDNA using indirect means to date, Piganeau and Eyre-Walker (2004) used five tests on six population data sets and obtained twenty-eight p values. Of these twenty-eight, seven (25%) were statistically significant. However, if we remove p values obtained using LD versus distance using D’ and the Max Chi Squared test (the implementation of Max Chi Squared in their study did not account for expected Chi square values of less than 2), seven out of seventeen p values achieve statistical significance, or around 41%.

Recombination in human mtDNA, if it is occurring, could not only have a dramatic impact on our understanding of disease etiology and transmission, but also on numerous evolutionary inferences. Even small amounts could adversely compound evolutionary dating and bias phylogenetic analyses (Schierup and Hein 2000a; b; White et al. 2008). Recombination can lead to false discovery of positive selection using the dN/dS ratio, leading to erroneous interpretation of the impact of selection across genomes (Anisimova et al. 2003; Arenas and Posada 2010a). Further, recombination has a direct and confounding effect on ancestral sequence reconstruction (Arenas and Posada 2010b). It is important, therefore, to obtain accurate estimates of mitochondrial recombination at the population level. Although attempts have been made, indirect evidence for mtDNA recombination in humans has been unconvincing thus far, in spite of compelling direct evidence. Here, we show
that even if recombination in human mtDNA were to occur, its detection is not guaranteed using the currently available tests, and if we were to project from the trends presented here we would expect values of rho to need to be around 3–4 times that seen in vivo (i.e. around 20), for rates of recombination detection to be in the 90% range for the tests reviewed here. This leads to the possibility that recombination could be occurring in human mtDNA at the population level, but that these levels are below current detection thresholds. This study is timely, as the advance of the new wave of sequencing technology will lead to many further opportunities to screen human populations for the presence of mitochondrial recombination. We foresee results from our work leading to improved accuracy of detection of human mitochondrial recombination frequency at the population level.

Acknowledgements

The authors would like to thank Konstantin Khrapko (Yale) for the original recombinant mtDNA dataset, Vladimir Mencl (Canterbury) for his programming expertise, Coen Willemse for support with the C translation of the Homoplasy Test, and David Posada (Vigo) for providing the C translation of the Homoplasy Test as well as making comments valuable to this manuscript. The manuscript was further enriched after useful comments from an anonymous reviewer. This work was funded by the Royal Society of New Zealand, Marsden Grant UOC402, and by start-up funds provided by the University of Otago to NJG.


Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia
Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas

Jakobsen, I. B., and S. Easteal, 1996 A program for calculating and displaying
compatibility matrices as an aid in determining reticulate evolution in

Jorde, L. B., and M. Bamshad, 2000 Questioning Evidence for Recombination in
Human Mitochondrial DNA. Science 288: 1931a.

Kivisild, T., and R. Villems, 2000 Questioning Evidence for Recombination in
Human Mitochondrial DNA. Science 288: 1931a.

Kraytsberg, Y., M. Schwartz, T. A. Brown, K. Ebralidse, W. S. Kunz et al., 2004

Kuhner, M. K., 2006 LAMARC 2.0: maximum likelihood and Bayesian estimation of

Kumar, S., P. Hedrick and T. Dowling, 2000 Questioning Evidence for
Recombination in Human Mitochondrial DNA. Science 288: 1931a.

Lakshmipathy, U., and C. Campbell, 1999 The human DNA ligase III gene encodes
nuclear and mitochondrial proteins. Molecular and Cellular Biology 19: 3869-
3876.

McVean, G., P. Awadalla and P. Fearnhead, 2002 A coalescent-based method for
detecting and estimating recombination from gene sequences. Genetics 160:
1231-1241.
Meunier, J., and A. Eyre-Walker, 2001 The correlation between linkage
disequilibrium and distance: Implications for recombination in hominid

selection shaped regional mtDNA variation in humans. Proceedings of the
National Academy of Sciences of the United States of America 100: 171-176.

Piganeau, G., and A. Eyre-Walker, 2004 A reanalysis of the indirect evidence for

Piganeau, G., M. Gardner and A. Eyre-Walker, 2004 A broad survey of
recombination in animal mitochondria. Molecular Biology and Evolution 21:
2319-2325.

Pluzhnikov, A., A. Di Rienzo and R. R. Hudson, 2002 Inferences about human
demography based on multilocus analyses of Noncoding sequences. Genetics
161: 1209-1218.

recombination from DNA sequences: Computer simulations. Proceedings of
the National Academy of Sciences of the United States of America 98: 13757-
13762.

Schierup, M. H., and J. Hein, 2000a Consequences of recombination on traditional

Schierup, M. H., and J. Hein, 2000b Recombination and the molecular clock.
Molecular Biology and Evolution 17: 1578-1579.

Schwartz, M., and J. Vissing, 2002 Paternal inheritance of mtDNA in a patient with


TABLE 1. Comparison of recombination detection rates for mtDNA between a simple model of evolution and a model that includes mutation rate heterogeneity, but no population growth.

<table>
<thead>
<tr>
<th>Test</th>
<th>Model</th>
<th>Recombination Parameter $\rho$</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>D'</td>
<td>simple</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>rh</td>
<td>5</td>
<td>75</td>
</tr>
<tr>
<td>$r^2$</td>
<td>simple</td>
<td>8</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>rh</td>
<td>3</td>
<td>71</td>
</tr>
<tr>
<td>HT</td>
<td>simple</td>
<td>30</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>rh</td>
<td>42</td>
<td>96</td>
</tr>
<tr>
<td>HT 3rd</td>
<td>simple</td>
<td>25</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>rh</td>
<td>39</td>
<td>95</td>
</tr>
<tr>
<td>PHI p</td>
<td>simple</td>
<td>0</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>rh</td>
<td>1</td>
<td>61</td>
</tr>
<tr>
<td>PHI n</td>
<td>simple</td>
<td>0</td>
<td>64</td>
</tr>
<tr>
<td></td>
<td>rh</td>
<td>1</td>
<td>60</td>
</tr>
<tr>
<td>NSS</td>
<td>simple</td>
<td>1</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>rh</td>
<td>3</td>
<td>71</td>
</tr>
<tr>
<td>MS</td>
<td>simple</td>
<td>8</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>rh</td>
<td>3</td>
<td>58</td>
</tr>
</tbody>
</table>

Values represent the number of 100 samples in which recombination was detected, using a nominal p value of 0.05 to indicate significance. P values are from paired, two-tailed student’s t-tests. Simple: simple scenario of sequence evolution; rh: evolution of sequences involves mutation rate heterogeneity, but no population growth; D': LD versus distance using D'; $r^2$: LD versus distance using $r^2$; HT: Homoplasy Test; HT 3rd: Homoplasy Test using only every 3rd site; PHI p: Pairwise Homoplasy Index using permutations; PHI n: Pairwise Homoplasy Index using a normal approximation; NSS: Neighbourhood Similarity Score; MS: Max Chi Squared.
FIGURE 1
Power of six indirect recombination tests at detecting recombination in simulated mtDNA sequence using (A) a simple mode of evolution, and (B) a slightly more complex mode of evolution that incorporates mutation rate heterogeneity and population expansion, over a range of recombination values. $D'$: LD versus distance using $D'$; $r^2$: LD versus distance using $r^2$; HT: Homoplasy Test; HT 3rd: Homoplasy Test using only every 3rd site; PHI p: Pairwise Homoplasy Index using permutations; PHI n: Pairwise Homoplasy Index using a normal approximation; NSS: Neighbourhood Similarity Score; MS: Max Chi Squared.

A.

% of simulated datasets that generated a p value of 0.05 or less for the simple scenario
B.

% of simulated datasets that generated a p value of 0.05 or less for the complex scenario

![Graph showing the percentage of simulated datasets that generated a p value of 0.05 or less for the complex scenario as a function of the population recombination parameter, \( \rho \). The graph includes various markers and lines for different statistics: \( D' \), \( r^2 \), HT, HT 3rd, PHI_p, PHI_n, NSS, and MS. The x-axis represents the population recombination parameter, \( \rho \), ranging from 0 to 14, and the y-axis represents the percentage, ranging from 0 to 100.]
FIGURE 2
P value frequency distribution at $\rho=0$ across the six indirect tests for the (A) simple model of sequence evolution (no rate heterogeneity or population growth) and (B) slightly more complex model of sequence evolution that incorporates mutation rate heterogeneity and population expansion.

A.