

The Original paper was published in *Heredity* 115: 396-404; advance online publication, May 6, 2015; doi:10.1038/hdy.2015.36

Three hundred years of low non-paternity in a human population

Jaco M. Greeff* & J. Christoff Erasmus

Department of Genetics, University of Pretoria, Pretoria 0002, South Africa.

* e-mail: jaco.greeff@up.ac.za

Abstract

When cuckoldry is frequent we can expect fathers to withhold investment in offspring that may not be theirs. Human paternal investment can be substantial and is in line with observations from tens of thousands of conceptions that suggest that cuckoldry is rare in humans. The generality of this claim seems to be in question as the rate of cuckoldry varies across populations and studies have mostly been on Western populations. Two additional factors complicate our conclusions, 1) current estimates of the rate of cuckoldry in humans may not reflect our past behaviour as adultery can be concealed by the use of contraceptives; and 2) it is difficult to obtain samples that are random with respect to their paternity certainty. Studies that combine genealogies with Y-chromosome haplotyping are able to circumvent some of these problems by probing into humans' historical behaviour. Here we use this approach to investigate 1273 conceptions over a period of 330 years in 24 families of the Afrikaner population in South Africa. We use haplotype frequency and diversity and coalescent simulations to show that the male population did not undergo a severe bottleneck and that paternity exclusion rates are high for this population. The rate of cuckoldry in this Western population was 0.9% (95%CI 0.4 – 1.5%) and we argue that given the current data on historical populations we have to conclude that, at least for Western human populations, cuckoldry rate is probably in the range of 1%.

Keywords: non-paternity; cuckoldry; extra-pair paternity; human; historical; genealogy

Introduction

Paternity certainty is an important parameter that affects the evolution of paternal behaviour (Trivers, 1972; Gaulin and Schlegel, 1980; Buss, 1996; Goetz and Shackelford, 2009a) and mating systems (Alexander *et al.*, 1979; Kokko and Jennions, 2008). Estimates of cuckoldry in humans must be one of the most accurate for any species with recent reviews summarising data for more than 24,000 conceptions (Bellis *et al.*, 2005; Anderson, 2006; Voracek *et al.*, 2008). We use the term conception throughout to indicate a conceived child that was born live. Despite the enormous sample size, early sentiments that there may be no standard human cuckoldry rate (MacIntyre and Sooman, 1991) have been echoed in subsequent studies where cuckoldry has been shown to vary across cultures (Bellis *et al.*, 2005), time (Voracek *et al.*, 2008) and between families (Anderson, 2006; Scelza, 2011).

Perhaps the most important finding of recent reviews of predominantly Western populations is that cuckoldry is normally very infrequent – in the range of 1.9% (Anderson, 2006) to 3.1% (Voracek *et al.*, 2008). Five subsequent studies on dissimilar populations also gave very low estimates of 0.94% in Germany (Wolf *et al.*, 2012), 0.73% for Afrikaners (Greeff *et al.*, 2012), 1.3%-2.9% for different Dogon religions (Strassmann *et al.*, 2012), 0.9% in Belgium (Larmuseau *et al.*, 2013), and 1.2% in Italy (Boattini *et al.* 2015). While this important parameter for human populations may not be pinned down to a single value, it is necessary to quantify the rate of cuckoldry in a variety of human populations and

sample sizes need to be large since small samples can overestimate the non-paternity rate (Voracek *et al.*, 2008).

Several factors have been shown to affect the rate of cuckoldry: cases where paternity was contested have at least an order of magnitude higher cuckoldry rate than estimates based on organ donation (29.8% versus 1.7%; Anderson, 2006). The former is biased for low paternity certainty whereas the latter may be enriched for men with high paternity confidence and could be misleadingly low (Anderson, 2006). Religious beliefs can keep cuckoldry very low (Alexander *et al.*, 1979; <3% Strassmann *et al.*, 2012). Even so, the more secular modern Western populations show a drop in cuckoldry of 0.8% per decade (Voracek *et al.*, 2008). Simmons *et al.* (2004) and Voracek *et al.* (2008) argued that such a drop can be the result of the use of modern contraceptives, especially the pill that has been introduced since the 1960s. If most modern-day extramarital affairs are concealed by contraceptives, studies on extant populations can give us a biased estimate of humans' "natural" cuckoldry rate (Larmuseau *et al.*, 2013).

A number of studies have combined deep-rooting genealogies with Y-chromosome typing to estimate cuckoldry rates (Heyer *et al.*, 1997; Foster *et al.*, 1998; Sykes and Irven, 2000; Soodyall *et al.* 2003; Kayser *et al.*, 2007; Pollin *et al.*, 2008; King and Jobling, 2009a; Strassman *et al.*, 2012; Greeff *et al.*, 2012; Larmuseau *et al.*, 2013; Boattini *et al.*, 2015). The advantage of these studies is that most of the conceptions occurred before the invention of contraception and apart from the sampled individuals, all the conceptions of earlier

generations should be random with respect to paternity certainty as most men already passed away. When a social father is excluded as a genetic father it may be due to cuckoldry or adoption. Genetic studies can frequently not distinguish between these alternatives and the term non-paternity is used to incorporate both these causes of a genetic mismatch between father and offspring. Even so, these estimates frequently give non-paternity rates below 1% (Greeff *et al.*, 2012). Cuckoldry by males carrying the same Y-chromosome cannot be detected in such historical studies (Foster *et al.*, 1998) and it needs to be confirmed that haplotype diversity is high enough to allow high exclusion rates of randomly drawn males.

C.C. De Villiers (1850-1887) worked on the baptismal and wedding records of the church archives to compile the complete genealogy of the Afrikaner population from 1652 – 1806. This work was published posthumously in 1893-94 as “Geslacht-register der oude Kaapsche familien” (Genealogies of old Cape families). These were improved and corrected by Pama and published in three volumes in 1966 as “Genealogies of Old South African Families” and at this time it included the genealogies of 2078 surnames and stretched into 1820. Based on the research notes of Heese (1971) and their own work the Human Sciences Research Council (of South Africa) and later the Genealogical Institute of South Africa (GISA) updated this work in a 17 volume series entitled “South African Genealogies”. This work included more recent immigrants and corrected previous mistakes and contained genealogies for over 8,000 families on its more than 10,000 A4 pages. This gargantuan task is still continuing with new research extending

families into the twentieth century and in 2014 GISA released the newest three of 23 CDs entitled “South African Family Registers” containing genealogies of 5,820 surnames running from a – l in more than 657,000 lines of text (www.gisa.org.za). Genealogies were thus not restricted to a few rich families. Using this advantage Greeff *et al.* (2012) showed with a sample of 25 men representing 111 conceptions on a deep-rooting genealogy that non-paternity in one Afrikaner family was about 0.9% (95% CI: 0.02% - 4.92%). This is a small sample size to estimate such low frequencies accurately with an upper 95% confidence limit of almost 5% and this family may not reflect the general population. Similar to other populations, contested paternities have a high cuckoldry rate of 22% (du Toit *et al.*, 1989).

Here we obtain a more accurate and representative sample of non-paternity in the Afrikaner population by investigating 1273 conceptions spread over twenty four families over the last three hundred and thirty years. While this population is at the southern tip of Africa it has a typical Western culture. By using deep-rooted pedigrees we can avoid biased sampling and the effects of contraception.

Materials and Methods

Afrikaner population

The Afrikaner population is mostly of European descent and began 360 years ago when Jan van Riebeeck of the East Indian Company founded a refreshment company at the Cape of Good Hope (Giliomee, 2003). The European component was roughly a third Dutch, a third German and a third French (de Villiers and

Pama, 1966; Heese, 1971; de Bruyn, 1976; Greeff, 2007). The French were Huguenots that fled France following the revocation of the Edict of Nantes in 1685 and were therefore ardent Protestants. The Dutch were also strongly religious and the Company had weekly church services. The German immigrants were mostly unskilled and financially marginalized immigrants that worked as sailors or soldiers. While the Afrikaner is a textbook example of a founder effect (Ridley, 2004), at least 5000 male immigrants joined the population over the first 200 years and claims of a founder effect are simply not true for the Y-chromosome (Table 1; Heese, 1971; see exclusion probability calculations below).

Table 1 Population dynamics of male Afrikaners for the first 240 years.

Time period	Immigrants ^a	Population size ^b
1657-1686	68	70
1687-1716	298	426.5
1717-1746	317	770.5
1747-1776	553	1789
1777-1806	758	3798
1807-1836	970	13008
1837-1866	2090	33493
1867-1896		86232

^a We extracted the number of male immigrants that came to the Cape in the first seven 30 year periods (210 years from 1657-1866) from Heese (1971); ^b We extracted the total adult male population size for every year from 1657-1819 from Gouws (1981). We used the average of the annual growth rate between 1747 and 1819 ($=1.032$, $SD = 0.024$) to extrapolate the population sizes between 1820 and 1866. The population size of each 30 year generation was then taken as the average of the 15th and 16th years of that 30 year stretch.

Ethics

Ethical clearance for this study was obtained from the ethics committee of the Faculty of Natural and Agricultural Sciences of the University of Pretoria (EC11912-065). The aims of the study were explained to all participants, participation was voluntary and all participants signed an informed consent form. They gave the details of their parents and grandparents to verify our genealogical information and to test for close relatedness via maternal lines.

Sample selection

In addition to the three genealogies that have been described (Greeff *et al.*, 2012; Greeff and Erasmus, 2013), we identified a further 20 old families that are well researched genealogically. The criterion for inclusion was simply that the family had to be well researched up until modern day. Note that in order to be inclusive the previously studied genealogies were included here. With the help of genealogists specialising on each surname we selected descendants of deep-rooting branches of each family (Figure 1), trying to balance the sample for immigrants from the Netherlands (9 families, 503 conceptions), Germany (8 families, 504 conceptions), France (5 families, 210 conceptions) and Scandinavia (1 family, 56 conceptions; Supplementary Table 1). The men were contacted and those who were willing to participate were sent an Oragene saliva self-collection kit (DNAgenotek, Canada) or came to the laboratory to give a saliva sample. In addition to the 25 Greeff (Greeff *et al.*, 2012) and 9 Botha (Greeff and Erasmus, 2013) males mentioned above, we obtained a further 165 that could be linked to known genealogies.

These 199 samples allowed us to study 1273 fertilizations. To quantify haplotype diversity further, we sampled a further 65 unlinked men from these selected surnames as well as another 28 samples from 21 different surnames.

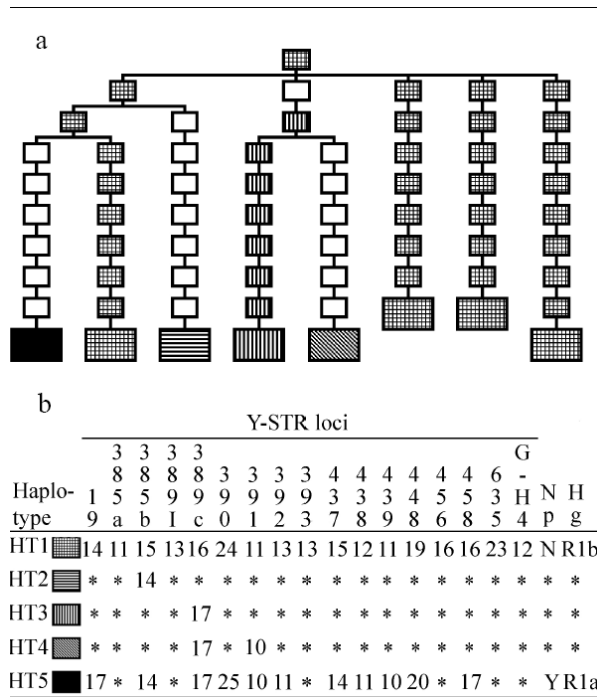


Figure 1 a) One example of the 24 genealogies compiled in this study. Genotyped individuals are indicated in larger rectangles, whereas inferred haplotypes are the smaller squares. Mutations are indicated by different patterns. Open squares indicate stretches of patriline where either mutations or non-paternities occurred. b) The number of repeats at each of 17 loci for each variant observed in the surname. Np is for non-paternity (Yes or No), * indicates that the number of repeats are the same as Haplotype 1 and Hg stands for haplogroup.

Genotyping

Whole genome DNA was extracted from saliva samples according to the manufacturers' instructions (DNAgenotek, Canada). Seventeen

y-chromosome STR loci (DYS19, DYS385 (a & b), DYS389 (I & II), DYS390, DYS391, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635 and GATA-H4) were amplified with the AmpF ϕ STR ® Yfiler ™ kit (Applied Biosystems, Foster City, USA) and length variation at each locus was detected with a 3500xl Genetic Analyser (Applied Biosystems, Foster City, USA) at the Sequencing Facility of Natural and Agricultural Sciences, University of Pretoria. Final editing of STR loci length polymorphisms were done with GeneMarker HID V2.2.0 (SoftGenetics, <http://www.softgenetics.com>). These 17 loci have been shown to be very good to discern between unrelated European men (Vermeulen *et al.* 2009). To quantify the degree to which the molecular variation on the Y-chromosome is explained by the country of origin of each immigrant and by the surname of the immigrant, we followed McEvoy and Bradley (2006) and did an AMOVA (Arlequin ver 3.5.1.2 (Excoffier *et al.*, 2005)). We grouped the descendants of each immigrant together in a population. For four surnames we knew that more than one immigrant came to the country (Supplementary Table 1) and we treat them separately here.

Inferring ancestral haplotypes

Given the importance of these 17 loci for forensic work and their rigorous capture on the YHRD data base, variation and mutation of these seventeen loci are well understood. A recent study by Goedbloed *et al.* (2009) considered 1,730 DNA-confirmed father-son transmissions at all of 17 loci, totalling 29,792 meiotic events and found 84 mutations in total of which all were single step mutation except

one two-step mutation. The number of repeat losses was equal to the number of gains (Goedbloed *et al.* 2009). In a larger meta-analysis of 18 studies and their own Goedbloed *et al.* (2009) considered a total of 135,212 meiotic transfers with most loci seeing more than 10,000 transfers and here they observed 331 mutational events with repeat gains outweighing losses although one has to be heedful of the allele length as shorter alleles are more likely to increase in length (Erasmus and Greeff in prep). Ninety seven point three percent of these mutations were single-step mutations with just over 2% being multi-step mutations. This set of microsatellites mutate at an average rate of 0.0028 (Goedbloed *et al.*, 2009) suggesting that we should see roughly 61 mutations (= 1273 conceptions x 17 loci x 0.0028 mutations per locus per fertilization) over all our trees. The greater majority of these (close to 99%) should also be single step mutations (Goedbloed *et al.*, 2009). To infer the haplotypes of the founders we worked backwards in time starting with our genotyped individuals and inferred the minimum number of mutations required that will result in the observed genotypes (Figure 1). Most of the time it was not possible to pinpoint mutations to one specific conception, rather it can be attributed to a stretch of patriline (Figure 1). Here we will use these data to determine the rate of non-paternity in the Afrikaner population. Once the non-paternities are excluded we will report on the mutation rates and nature of the mutations elsewhere.

Identification of non-paternity patrilines

The more generations an individual is removed from a founder, the higher the chances that

there will be mutations at one or a number of STR loci. But where do we draw the line and infer a non-paternity rather than several mutations. Kayser and Sajantila (2001) showed that for 15 of these 17 loci in 4,999 confirmed conceptions, at most two loci mutated. Three or more mutations were thus considered a non-paternity rather than mutations that occurred in one generation. On the other hand Goedbloed *et al.* (2009) found one confirmed father-son pair from 1,764 that had three mutations at these loci and suggested that locus specific mutation rates should be considered rather than a simple more than three rule for instance. In the present study we frequently deal with longer stretches of patriline and we need to obtain a yardstick for several generations. In order to do so we did Monte Carlo forward simulations of patrilines that vary in length from 1 to 10 and we used the locus specific mutation rates given in Goedbloed *et al.* (2009; their Table 2 that is mostly based on 10,000 conceptions excepting a few loci where estimates are based on roughly 3,500 conceptions). In the simulations we assumed that mutated loci got longer and shorter with equal probability (Rolf *et al.*, 2001; Goedbloed *et al.*, 2009). Seeing that the allele distributions of these loci show no skewness on the YHRD data base (Willuweit and Roewer, 2007) considering tens of thousands of samples, it is highly unlikely that lengthening of alleles outweigh shortening at any of these loci. We further assumed that mutations are normally single steps but that at a rate of 1.2 per 100 they lengthen or shorten by two repeat units (Goedbloed *et al.*, 2009 based on 135,212 meiotic transfers). We ran a million simulations for each patriline length, each time recording

the number of mutations (Table 2). All simulations and statistics were done in R (R Development Core Team, 2010). We counted loci that had two one-step mutations in the same direction and those that had a single two-step mutation as two mutations so that it is comparable to our own data where we cannot distinguish between these two mutational events. Even when a specific event occurs very

infrequently, in a big study like this with 1273 conceptions one may expect to see rare events. For this reason, we also calculated the number of times we expect to see an event given that the entire 1273 conceptions can result in 141 stretches of nine fertilizations, 159 of eight etc. (Table 2).

Table 2 Percentage and number of patriline of length equal to generation number that contain zero, one, two, three and four or more mutations.

Number of generations	Number of mutations ^a					1273 ÷ generations	Number of mutations ^b				
	0	1	2	3	>3		0	1	2	3	>3
1	95.49	4.36	0.14	0.00	0.00	1273	1215.55	55.56	1.84	0.04	0.00
2	91.17	8.35	0.46	0.02	0.00	637	580.30	53.16	2.91	0.12	0.00
3	87.04	11.98	0.93	0.05	0.00	424	369.33	50.85	3.93	0.22	0.01
4	83.18	15.18	1.52	0.10	0.01	318	264.73	48.32	4.85	0.33	0.02
5	79.35	18.25	2.20	0.18	0.01	255	202.02	46.47	5.61	0.47	0.03
6	75.92	20.84	2.94	0.28	0.02	212	161.07	44.21	6.23	0.60	0.05
7	72.47	23.21	3.85	0.43	0.04	182	131.80	42.21	7.00	0.78	0.07
8	69.21	25.40	4.73	0.60	0.06	159	110.13	40.42	7.53	0.95	0.10
9	66.11	27.28	5.71	0.80	0.10	141	93.51	38.59	8.07	1.14	0.14
10	63.22	28.90	6.69	1.05	0.14	127	80.48	36.79	8.52	1.33	0.17

^aThe left hand side gives the expected percentage of patriline of length equal to generation number that are expected to have zero, one, two, three and four or more mutations; ^b The right hand side gives the product of these percentages/100 and (1273/generation number).

Based on Table 2 we excluded all cases where 3 or more mutations occurred as non-paternities. In the identified cases of non-paternity we also counted the number of loci where two- or more-step mutations were required. All males that were excluded as descendants from their paternal ancestor were genotyped again to confirm that no laboratory errors occurred. Haplogroups generally precedes the use of surnames (King and Jobling, 2009b) and these were estimated for each haplotype with the Whit Athey haplogroup

predictor (Athey, 2005, 2006, <http://www.hprg.com/hapest5/>).

Exclusion probabilities

We know that the first two generations were respectively founded by 68 and an additional 298 male immigrants (Table 1; Heese 1971) and these relatively low numbers could potentially result in low paternity exclusion probabilities. The paternity exclusion probability/rate is the chance of excluding a random male as a father based on DNA evidence. Hence, if a few early immigrants had many male descendants, these descendants,

bearing the same surname, could cuckold each other without leaving any Y-DNA trace because their Y-haplotypes are identical by descent. Alternatively, cuckoldry may go undetected if the descendants of different immigrants carry identical or very similar Y-haplotypes. We deal with this potential pitfall below, but first we consider the demographic exclusion rate per se.

Note that the exclusion probability is in essence the chance that two different samples drawn at random from a specific generation of a specific population are not identical by descent. In other words, samples that have identical haplotypes if they coalesce within the Afrikaner population cannot be excluded. To calculate this probability we ran backward coalescent simulations for the first 240 years (or eight generations) written in R (R Development Core Team, 2010). We choose this time interval because we have very good demographic information for this period (Table 1; Heese, 1971; Gouws, 1981) which covers the foundation years of the colony. Furthermore, a founder effect resulting in one male's descendants becoming very dominant, and subsequently resulting in low exclusion rates, must occur early in the population's history.

We calculated an exclusion rate for each generation as it can vary depending on the population's demography. For each generation's estimate we repeated the following process 1 million times and calculated the fraction of times haplotypes did not coalesce. A male ancestor for each male was randomly drawn from the population size given in Table 1 for the previous generation. Sampling male ancestors from the previous generation

continued until either the same male was drawn (a coalescence event occurred) or when an ancestor was an immigrant (coalescence within the Afrikaner population was not possible). Here we thus assume that all immigrants had a unique haplotype and we will deal with this caveat below. Typical of backward coalescence simulations, the random sampling of males ensures a Poisson distribution of offspring so that some patrilineages can grow while others go extinct.

Now we will consider the chances that two immigrants introduced identical or very similar Y-haplotypes. First, the Y-filer kit is known to distinguish very well between unrelated European men (Vermeulen *et al.*, 2009). Yet, Larmuseau *et al.* (2014) sampled from a tiny part of Europe (compared to where Afrikaner progenitors originated from) and found that in a sample of 1028 men, 98 pairs had identical Y-haplotypes using the AmplF1STR Yfiler. Below we incorporate their measures of identity into our exclusion probability. Second, we directly tested how common Afrikaner haplotypes are in Europe from the YHRD site (Willuweit and Roewer, 2007). The probability of sampling two identical haplotypes given their European-specific frequencies is simply $\sum p_i^2$, with the summation being over all i haplotypes. The chance that two randomly sampled haplotypes are not identical is one minus the summation and in turn it can be multiplied with the demographic estimate to obtain a corrected exclusion probability.

Third, we determined how different haplotypes are from one another by making pairwise comparisons between all haplotypes, each time recording the number of loci that differ and the number of mutational steps

between haplotypes. We did it for a) the 26 founding haplotypes we inferred from our data and b) for the 26 founders, plus identified non-paternities plus the additional surnames we sampled (60 in total).

Fourth, to get an appreciation for how haplotype similarity and homoplasmy may affect exclusion probabilities we calculated the chance that two haplotypes that differ by one mutational step at each of n loci will mutate to be identical over x generations. For loci that are already the same and must not mutate, the probability is given by $(1 - \mu)^{2x}$, where μ is the locus specific mutation rate and the 2 in the power is because the locus must not mutate in both founder lineages. For lineages that have to mutate to become similar the probability is given by $x\mu(1 - \mu)^{2x-1}$. Note that only half the mutations will bring differing loci closer to each other, but at the same time, either of the lineages could have a mutation cancelling out the former effect. The probability is multiplied by x because there are x occasions where the mutation can occur. To obtain the probability over all 17 loci the probabilities of each are multiplied with one another.

Fifth, 33 of the 60 haplotypes fall in haplogroup R1b (see Results) as can be expected from a population founded by European men (Wells *et al.*, 2001). This means that a random father may be harder to exclude from R1b families than from the other less frequent haplogroups. To determine if this is the case we counted and compared the non-paternity rates for R1b families and non-R1b families.

Results

Sample characteristics

The births of the 1273 boys were spaced out over a period of 335 years from 1674 with a mean at 1859 (Figure 2). Apart from the first three and last two 30 year slots all periods had more than 100 births (i.e. conceptions). Ninety three percent of these births preceded 1960.

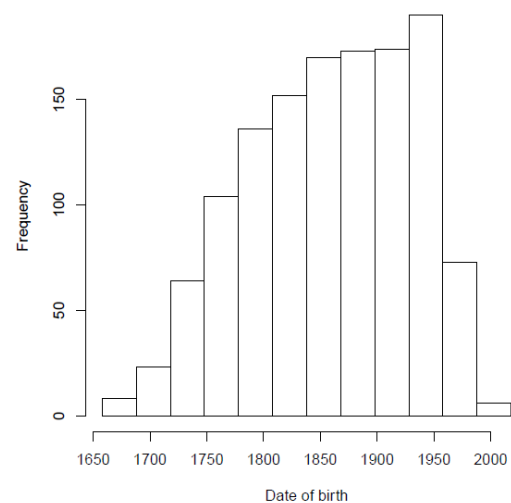


Figure 2 The date of birth of the 1273 boys included in the study.

The mean generation time, i.e. the father's age at the birth of his offspring was 32.92 years ($N = 1273$, $SD = 8.18$; Figure S1). Eight fathers were older than 60 and seven younger than 20. The age of fathers did not covary with the child's date of birth (correlation test: $t = -1.4038$, $df = 1271$, $P = 0.1606$, $\rho = -0.039$).

Inferring haplotypes

We observed 129 unique haplotypes in the study of which 69 were newly formed in the population by mutations (Supplementary Table 2). Country of origin explained none of the genetic variation whereas surname explained

82.2% of the genetic variation. Only 18.8% of the genetic variation was explained within families due to mutations and non-paternities. Males bearing the same surname normally had the same haplotype (bar mutations) and males with different surnames had different haplotypes. There was one exception where one non-paternity case had an identical haplotype to another surname.

Identification of excluded patriline

While we expect to see zero, one and even two mutations frequently in longer patriline, three mutations are very infrequent, even in the longest patriline (Table 2). Three or more mutations were thus considered a non-paternity. Based on this criterion we identified 11 non-paternity events in the 1273 father-to-son transmissions (Table 3). The identification of these eleven events is further strengthened by the fact that all of these cases required at least one, and frequently several multi-step mutations to explain the observed allele lengths (Table 3). The median non-paternity differed at nine loci from its proposed family and required a median of at least two multi-step mutations.

The estimated fraction of males excluded as their fathers' sons is thus 0.00864 ($=11/1273$) with 95% CI ranging from 0.00432 - 0.01541 (we make corrections for the probability of exclusion below). One family had two non-paternities that may involve one couple. The remaining families either had one or zero non-paternities. It is possible that we may have missed some non-paternities because some of the non-paternity branches may have witnessed more than one cuckoldry event.

Table 3 Characteristics of the 11 identified non-paternities.

Number of generations separating non-paternity from pedigree	Number of loci mutated	Number of loci with multi-step mutations
6	3	1
7	4	1
1	5	2
8	8	1
4	9	1
6	9	3
7	11	2
6	13	4
6	13	6
7	13	6
6	15	3

However, if we take into account that the total non-paternity branches sum to 64 father-to-son transfers, i.e. 5% of the total study and assuming a homogenous non-paternity rate of 0.00864 one expects to see only about 0.5 extra cuckoldries, then these branches can at most hide one more non-paternity event giving an estimate of 0.00942.

Exclusion probabilities

It is a misconception to assume that the male population went through a bottleneck (Table 1). Based on demography alone, the probability of excluding a random male as the father was 99.8686% in the first generation where cuckoldry was possible (Table 4). In subsequent generations it rose slightly (Table

4) presumably due to the addition of more male immigrants (Table 1) and a small pox epidemic in 1713 that killed an estimated 25% of the population (Gouws, 1981).

Of the 98 male pairs that had identical Yfiler haplotypes, 43 pairs were in fact related leaving 942 samples (=1028-(43x2)). Thus of the 443,211 possible pairs (=942x941/2), only 55 were identical. That is a probability of 1.240944×10^{-4} that two unrelated males (based on genealogies going back to 1800) have an identical haplotype with the Yfiler. Larmusseau *et al.* (2014) give the frequency of one step neighbours, which is much higher, but unfortunately do not give the frequency of

related males in this sample, which would also be much higher. Even so, with 519 one step pairs out of 527,878 (=1028*1027/2) possible pairs the probability of drawing two one step neighbours is 9.831817×10^{-4} . We can combine either of these probabilities with our demographic estimator by multiplying the estimator with one minus the fraction (Table 4) as that would be the chance that two sampled males are not identical and can thus be excluded. Note that even with the more conservative estimate, exclusion probabilities are still very close to 1 (Table 4).

Table 4 Exclusion probabilities for the first eight generations.

Generation	Demographic	YHRD frequencies ^b	Brabant identical ^c	Brabant 1 step ^c
1657-1686	^a			
1687-1716	0.998686	0.998669	0.998562	0.997704
1717-1746	0.998704	0.998687	0.998580	0.997722
1747-1776	0.998769	0.998752	0.998645	0.997787
1777-1806	0.998864	0.998847	0.998740	0.997882
1807-1836	0.998836	0.998819	0.998712	0.997854
1837-1866	0.998801	0.998784	0.998677	0.997819
1867-1896	0.998809	0.998792	0.998685	0.997827

^a Since one cannot cuckold yourself, the first generation does not have a cuckoldry estimate.

^b Demographic estimate corrected for upper 95% CI for our sampled haplotypes in $\sum p_i^2$, see text for detail.

^c Demographic estimate corrected for localised European sample from the areas of Flanders (Belgium), North Brabant and Limburg (Netherlands), see text for detail.

The frequencies of observed haplotypes in Afrikaners varied between 0 and 8.97×10^{-4} in Europe with a mean of 6.78×10^{-5} . Conservatively using the upper 95% CI of the haplotype frequency estimates (rather than the estimates themselves) the chance of sampling two identical haplotypes in the Afrikaner population based on their European frequencies is only 1.75×10^{-5} . This correction

is very small and all exclusion probabilities remain very close to 1 (Table 4).

Eleven of the families we investigated came from the first time period and 8 from the second and the remaining four came later. Therefore it is important to make this exclusion correction. However, because the paternity exclusion rate is so close to one (Table 4) and the non-paternity rate is so low (smaller than 0.01) the corrections have a small effect. If we

conservatively use an exclusion rate slightly lower than all our estimates (Table 4), 0.9977, to calculate the true non-paternity rate (Sasse *et al.*'s (1994) moment estimate) we obtain a very low non-paternity rate of 0.00866 (95% CI: 0.0043 - 0.0154). Note that even if our exclusion probability was as low as 0.87 (as compared to our 0.9977), we would still conclude that the true non-paternity rate was less than 1%.

Given that 82% of genetic variation is explained by surnames it is clear that our assumption that immigrants had unique haplotypes is well supported. The haplotypes were assigned to nine haplogroups with R1b being the most frequent (33 out of 60).

Next we consider the possibility that two immigrants had very similar haplotypes that subsequently mutated so that the exclusion probability is in fact less than we calculated.

When we compared all the inferred founders to each other they differed at 10 loci on average with an average of 15 single step mutations (Figure 3 a and b). Six of the 325 pairwise comparisons had three mutational steps between them and one had two mutational steps. These similar haplotypes all belonged to R1b. If we consider all 60 haplotypes identified the average pair differed at 10 loci and an average of 16 mutational steps (Figure 3 c and d). One of the 1770 (=0.00056) pairwise comparisons was identical and this was the previously mentioned non-paternity. Two haplotypes differed by one mutational step from each other (Table 5). In addition to the one two-step difference seen between the founder haplotypes three more were observed in this comparison and we found 20 comparisons with three step differences (Table 5).

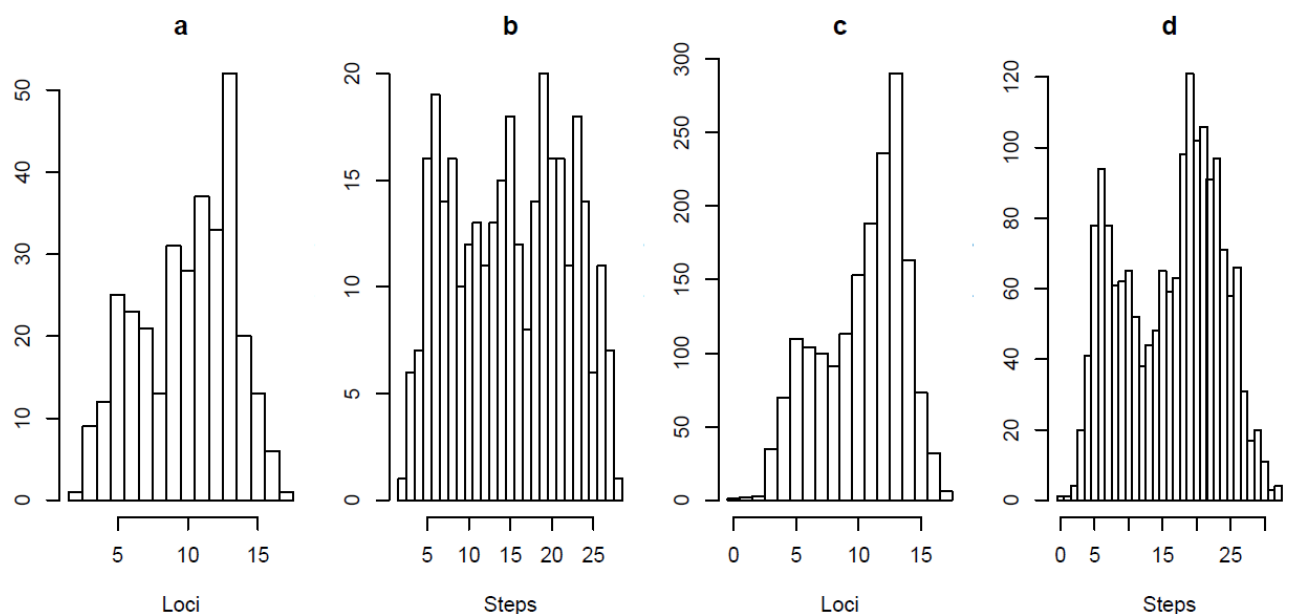


Figure 3 Histograms from pairwise comparisons of the 26 founders (a and b) and all 60 haplotypes (c and d). The number of loci that differ (a and c) and the number of steps (b and d) between each pairwise comparison are given.

Table 5 The chance that two haplotypes mutates to be identical over ten generations.

Number of mutational steps between haplotypes	Empirical occurrence ^a	Probability of mutating to be the same ^b	Combined probability of observing ^c
1	$\frac{1}{1770}$	0.0108	$\frac{6}{1,000,000}$
2	$\frac{4}{1770}$	0.0003	$\frac{7}{10,000,000}$
3	$\frac{20}{1770}$	8.4×10^{-6}	$\frac{9}{100,000,000}$

^a Obtained from 1770 pairwise comparisons; ^b Using equation 1 with average mutation rate; ^c the product of columns 2 and 3.

Working with the average mutation rate of 0.002776, the chances that two haplotypes differing by one mutational step mutate to be identical over a period of ten generations is 0.0108, taking into account that this occurred in only 1 out of 1770 comparisons it gives us a combined probability of 6 in a million (Table 5). We can work out similar probabilities for haplotypes differing at two and three loci (Table 5). Note that we did not use the locus-specific mutation rates as the loci affected differ between comparisons. To include the seldom observed similarity between haplotypes in the exclusion probability we need to realise that we have calculated the probability that someone previously thought to be excluded because immigrant haplotypes were unique can now not be excluded because haplotypes mutated to be similar. If we notice that the likelihood of this observation decreases exponentially as there are more mutational steps between immigrants (Table 5) it is improbable that the right-hand column of Table 5 will sum to more than 7 in a million. Therefore the exclusion probability

should be $0.998686 - 0.000007 = 0.998679$ rather than 0.998686. This correction is thus miniscule.

We observed 730 conceptions in R1b families and 543 in non-R1b families. In the former we identified 8 non-paternities whereas the latter only had 3. Non-paternities were thus in fact less frequent in non-R1b families than R1b families, although not significantly so (Non-paternity rates: R1b families = 0.011, non-R1b families = 0.006; $\chi^2 = 0.533$, $df = 1$, $P = 0.466$). The inclusion of R1b families does not bias our non-paternity estimate downwards.

Discussion

In this study of 1273 conceptions over a period of 335 years we find a very low non-paternity rate of below 1% and most probably not exceeding 1.5%. Even with very low exclusion rates of 0.87 (as compared to our 0.9977) we still calculate a non-paternity rate of below 1% per generation. Since we used deep-rooting genealogies 92% of our investigated conceptions were not from sampled males themselves but from their ancestors. In this way we, and other studies relying on pedigrees,

have circumvented to a large degree the problem of biased sampling with respect to paternity certainty. Furthermore, 93% of our studied conceptions predate the invention of the pill and contraception could not have concealed extramarital affairs. This combination of genealogies with genetic data give us a novel understanding of human historic behaviour (Foster *et al.*, 1998; Sykes and Irwen, 2000; Jobling, 2001; Soodyall *et al.* 2003; Kayser *et al.*, 2007; Larmuseau *et al.*, 2013; Boattini *et al.*, 2015).

Despite the fact that the Afrikaner population is believed to have had a strong founder effect (Ridley, 2004), it is certainly not the case for the male population. As many as 5000 male immigrants were recorded (Table 1; Heese 1971) leading to exclusion probabilities in excess of 99.86% (Table 4). Incorporating the probability that immigrants were not unique based on European allele frequencies lead to a small adjustment of the exclusion probability to 99.77%. Furthermore, we found a high haplotype diversity between founders (Figure 3) using the Y-filer kit that is very good at discriminating between unrelated European males (Vermeulen *et al.*, 2009). The high variation supports the observation that many immigrants (Table 1; Heese, 1971) carrying unique haplotypes (Figure 3) came to the Cape. We further showed that in a few cases where haplotypes were similar enough that they could conceivably mutate to be the same these were so infrequent and improbable that the adjustment to the exclusion probability is in the order of 7×10^{-6} (Table 5). Finally, the non-paternity rate of R1b families was higher, rather than lower, than non-R1b families, suggesting

that we did not miss non-paternities in the more frequent R1b families.

As more studies are reporting very low non-paternity rates (Anderson, 2006; Voracek *et al.*, 2008; Strassmann *et al.*, 2012; Wolf *et al.*, 2012; Larmuseau *et al.* 2013; Boattini *et al.*, 2015) in disparate human populations we may need to conclude that the historical cuckoldry rate of, at least Western, human populations may well be very low – less than 3% and frequently less than 1%. In support of such low rates of non-paternity several studies comparing surnames to Y-chromosome haplotypes of substantially older populations also gave very low rates of non-paternity: 1.3% for Sykes in England (Sykes and Irwen, 2000); 1.6% for O'Sullivan in Ireland (McEvoy and Bradley, 2006); 1.49% in Iceland (Helgason *et al.*, 2003); 0.74% for the 5 most common surnames in Oriente Columbia (Bedoya *et al.*, 2006); 1.28%-3.26% for 5 British surnames (King and Jobling, 2009a); 0.91% in the Belgian population (Larmuseau *et al.*, 2013); 1.21% in the *Partecipanza* of Italy (Boattini *et al.*, 2015). A further three studies using deep-rooting genealogies also recorded an average rate of below 1% (Greeff *et al.*, 2012).

It is clear that there is no single human mating strategy (Buss, 2007; Wlodarski *et al.*, 2015). This is perhaps chiefly due to an asymmetry in the desires of the two sexes (Trivers, 1972; Buss, 1989; Buss and Schmitt, 1993), but also due to the asymmetry in information; namely that females can be certain of their maternity, whereas males cannot be sure of their paternity (Trivers, 1972; Goetz and Shackelford, 2009a). As a result, the quantification of non-paternity can give us some information on the mating system, but

does not define the mating system (Gomendio *et al.*, 1998). For instance, the low non-paternity observed here and in other studies may suggest high female partner fidelity in humans, but it may also reflect low partner fidelity in females combined with well-developed anti-cuckoldry tactics of males. We will consider these alternatives in turn.

Giliomee (2010) pointed out that in the settlements' early days women were more likely to be churchgoing than men. The Protestant religion is indeed still a corner stone of the Afrikaner culture (Giliomee, 2003) and the explicit prohibition of adultery in the Ten Commandments may well be the best explanation for the low rate of non-paternity in this population and possibly others (Alexander *et al.*, 1979). In the Dogon, Christians had a non-paternity rate of 2.9% (Strassmann *et al.*, 2012). But Christianity is still very new to the Dogon and intertwined with traditional beliefs which can explain its lower efficacy at reducing cuckoldry (Strassmann *et al.*, 2012). The slave lodge at the Cape served as a brothel for an hour every day (Giliomee and Mbenga, 2007; Groenewald, 2007). Children born from these premarital and potentially extra-marital "affairs" of men would default to the slave population (Groenewald, 2007). For many christenings of slave-born children fathers are not recorded - presumably because European men would not own up to their paternity or because a multitude of men may have been the father. Genetically this can also be seen from the significant European male contribution to the Cape Coloured population while an European female contribution is almost completely lacking (De Wit *et al.*, 2010; Quintana-Murci *et al.* 2010).

Non-paternity rate of Afrikaners may thus tell us more about women's chastity than men's.

On the other hand, a multitude of pre- and post-copulatory mechanisms exist to reduce cuckoldry (Smith, 1984). To name but a few: males' universally observed sexual jealousy (Daly *et al.* 1982), men's higher sexual interest in their partners following a period of absence (Shackelford *et al.*, 2002; Starratt *et al.*, 2007), the larger amount of sperm transferred after a period of absence (Baker and Bellis, 1993a), frequent inseminations by the primary partner (Baker and Bellis, 1993a; Goetz *et al.*, 2007) and forced in-pair copulations (Goetz and Shackelford, 2006, 2009b). This diverse array of male anti-cuckoldry practices that minimize the effect of female infidelity and conception risk strongly suggests that paternity uncertainty has been a longstanding adaptive problem for man (Buss, 1996; Shackelford *et al.*, 2005).

The fact that questionnaire based studies on females' behaviour and perceptions frequently calculate higher non-paternity rates (Bellis and Baker, 1990; Koehler and Chisholm, 2007; Voracek *et al.*, 2009; Scelza, 2011) suggests that females may overestimate their success in extra-marital affairs or at least confuse extra-pair copulations with cuckoldry (Simmons *et al.*, 2004). Another factor that can substantially reduce the effectiveness of extra-pair copulations in humans is humans' relatively low fecundability in combination with concealed ovulation (Alexander and Noonan, 1979; Gomendio *et al.*, 1998). This combination of traits will reduce cuckoldry and allow the steady partner high paternity confidence.

In line with expectations (Shackelford and LeBlanc, 2001) our data suggest that

cryptic female choice favouring extra-pair partners is very limited and/or unsuccessful in the Afrikaner population. Earlier studies suggested a potential role for cryptic female choice in humans (Smith, 1984; Baker and Bellis, 1993b), but we see very little support for it here.

The non-paternity rate of around 1% observed here and in many other studies of Western populations is not enough to surpass the paternity threshold to explain avuncular societies (Greene, 1978) but it may be sufficient (Parker, 1982) to explain considerable selection on male behaviour, psychology, physiology and morphology (Pound *et al.*, 2006; Shackelford *et al.*, 2005).

The approach followed here cannot detect cuckoldry by close male relatives (Foster *et al.* 1998) and may thus be an underestimate. However, it should be noted that the exclusion rates are very close to one. It is possible that sons that were the product of cuckoldry may have been identified as such and subsequently neglected so that they may have been less likely to leave descendants and hence less likely to be included in our study. While such incidents would mean that our and other pedigree based estimates of cuckoldry would be too low it also implies that in these cases men did not squander energy on cuckolded offspring. Since the cost of being cuckolded would not have been paid by such fathers it is arguable whether these cases should be included in an estimate of cuckoldry. On the other hand the cost of a wrongful identification of a cuckoldry event, especially given the suggested anonymity of babies, would be high (Pagel, 1997). Perhaps for this reason, divorces based on cuckoldry were very

infrequent with only three recorded in the first 50 years of the colony's existence (de Wet, 1981; Greeff and Erasmus, 2013). It is also the case that due to the low number of females at the Cape in the early days, some men lived together with former slaves or Khoikhoi women (Groenewald, 2007). These couples were dissuaded by law and were not allowed to marry (Groenewald, 2007) and would have been systematically excluded from the current sample. On the other hand, unknown cases of adoption will lead us to incorrectly infer a cuckoldry event. Specifically in the Afrikaner population teenage pregnancies out of wedlock were sometimes concealed by the girl's mother pretending to be pregnant and then raising her grandchild as if it was hers. These so-called grandparental adoptions (Larmuseau *et al.*, 2013) of boys will introduce a new Y lineage to the surname and suggest a cuckoldry event while none occurred. Given the very low estimate of non-paternity in this study, we can also conclude that the rate of clandestine grandparental adoptions must also have been very low.

The gene arginine vasopressin receptor 1a (*AVPR1a*) has been shown to affect happiness of men in monogamous relationships. Given the Afrikaners' founder effect amongst women, one potential explanation for their low non-paternity rate may have been a very different allele frequency at *AVPR1a* compared to other populations. However, we recently showed that its allele frequency is virtually identical to other European populations (Erasmus *et al.*, submitted).

For Irish families only 20% of variation was explained between families (McEvoy and

Bradley, 2006) while this is 82% in the Afrikaner population. Several factors can explain this difference. Due to the Afrikaner population being younger less time has elapsed for mutations and cuckoldry to scramble the genetic signal of the family name. In our study we targeted families with single founders and in three cases we knew more than one immigrant came to South Africa bearing the same surname and we treated them separately. In the British Isles common surnames tend to stem from many founding events and have several haplotypes, whereas smaller surnames are more likely to share a single haplotype (King and Jobling, 2009a,b). While our sample included the three most common Afrikaans surnames they were all uniform with respect to haplotypes. From records we knew they were founded by single founders and the families' sizes are a result of differential growth. These preliminary findings strongly suggest that Y-STR haplotypes can be used to infer a most likely surname list for forensic purposes for the Afrikaner population. More surnames will need to be sampled to confirm the uniqueness of Y-haplotypes of immigrants and to compile the haplotype lists.

While the average male started reproduction at an age of 27 years (Greeff and Koberstein, unpublished data) the average generation length was almost 33 years which is very close to 33.38 years observed by Boattini *et al.* (2015) for another historic population. Given that average generation time is frequently a pivotal assumption to models and data analyses it is important to remember that males' reproductive period stretched over a long time and that samples will include children irrespective of their birth order.

Currently, most, but not all studies on cuckoldry rates have been on typical Western populations. The most consistent finding is a very low cuckoldry rate. Studies like ours that look at historical data also indicate very low cuckoldry rates. These findings suggest that while selection on anti-cuckoldry tactics may be ubiquitous, it is low, and low cuckoldry rates may have coevolved with our high paternal care (Geary, 2006).

Acknowledgements

We thank Isabel Groesbeek for helping us select well studied Afrikaner families and putting us into contact with the relevant genealogists. We thank the genealogists, Richard Ball, Johan Barkhuizen, Joe Botha, Petrie Coetzee, Jaleen de Beer, Annatjie Erasmus, Danie Fourie, Louis Kamfer, Nantes Kruger, Hennie Labuschagne, Herman Labuschagne, Linda Lombard, Sasa Malan, Janet Melville, Ria Meyer, Freek Myburgh, Faan Naudé, Cobus Potgieter, Colin Pretorius, Paul Prinsloo, Peter Sibernagl, John Stephens, Chris Steyn, Adriaan Swanepoel, Piet Venter and Christo Viljoen who helped us identify descendants of deep branches of their families and put us into contact with them. We thank three reviewers for their thought-provoking comments. The running cost for this project was funded by the National Research Foundation (NRF) grants 65696, 72737 and 77256 to JMG. JCE was supported by a free-standing NRF scarce skills grant. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and therefore the NRF do not accept any liability in regard thereto.

Conflicts of Interest

Because this work was conducted to determine the non-paternity rate in humans, including both authors' families, every effort was made to give an unbiased opinion. The authors have no other interests to declare.

Data archiving

R code for coalescence simulation is given in supplementary material.

Haplotypes of linked and unlinked founders are given in Supplementary Table 2.

References

- Alexander RD, Noonan KM (1979). Concealment of ovulation, parental care, and human social evolution. In: Chagnon NA, Irons W (eds) *Evolutionary biology and human social behavior: An anthropological perspective*, Duxbury Press: Belmont. pp 436-453.
- Alexander RD, Hoogland JL, Howard RD, Noonan KM, Sherman PW (1979). Sexual dimorphisms and breeding systems in pinnipeds, ungulates, primates and humans. In: Chagnon NA, Irons W (eds) *Evolutionary biology and human social behavior: An anthropological perspective*, Duxbury Press: Belmont. pp 402-435.
- Anderson KG (2006). How well does paternity confidence match actual paternity? Evidence from worldwide nonpaternity rates. *Curr Anthropol* **47**: 513-520.
- Athey TW (2005). Haplogroup prediction from Y-STR values using an allele-frequency approach. *J Genet Geneal* **1**: 1-7.
- Athey TW (2006). Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *J Genet Geneal* **2**: 34-39.
- Baker RR, Bellis MA (1993a). Human sperm competition: ejaculate adjustment by males and the function of masturbation. *Anim Behav* **46**: 861-885.
- Baker RR, Bellis MA (1993b). Human sperm competition: ejaculate manipulation by females and a function for the female orgasm. *Anim Behav* **46**: 887-909.
- Bedoya G, Montoya P, García J, Soto I, Bourgeois S, Carvajal L, *et al.*, (2006). Admixture dynamics in Hispanics: A shift in the nuclear genetic ancestry of a South American population isolate. *Proc Natl Acad Sci* **103**: 7234-7239.
- Bellis MA, Baker RR (1990). Do females promote sperm competition? Data for humans. *Anim Behav* **40**: 997-999.
- Bellis MA, Hughes K, Hughes S, Ashton JR (2005). Measuring paternal discrepancy and its public health consequences. *J Epidemiol Community Health* **59**: 749-754.
- Boattini A, Sarno S, Pedrini P, Medoro C, Carta M, Tucci S, *et al.*, (2015). Traces of medieval migrations in a socially stratified population from northern Italy. Evidence from uniparental markers and deep-rooted pedigrees. *Heredity* **114**: 155-162.
- Buss DM (1989). Sex differences in human mate preferences: evolutionary hypotheses tested in 37 cultures. *Behav Brain Sci* **12**: 1-49.
- Buss DM (1996). Paternity uncertainty and the complex repertoire of human mating strategies. *Amer Psychol* **51**: 161-162.
- Buss DM (2007). The evolution of human mating. *Acta Psychol Sinica* **39**: 502-512.
- Buss DM, Schmitt DP (1993). Sexual strategies theory: an evolutionary perspective on human mating. *Psychol Rev* **100**: 204-232.
- Daly M, Wilson M, Weghorst SJ (1982). Male sexual jealousy. *Ethol Sociobiol* **3**: 11-27.

- De Bruyn GFC (1976). Die samestelling van die Afrikanervolk. *Tydskrif vir Geesteswetenskappe* **16**: 39-42.
- De Villiers CC, Pama C (1966). *Geslagsregisters van die ou Kaapse families, 1 A-K*. A.A. Balkeman: Kaapstad.
- De Wet GC (1981). *Die vryliede en vryswartes in die Kaapse nedersetting 1657-1707*, Historiese Publikasie Vereniging: Kaapstad.
- De Wit E, Delport W, Rugamika CE, Meintjes A, Möller M, Van Helden PD, *et al.*, (2010). Genome-wide analysis of the structure of the South African Coloured population in the western Cape. *Hum Genet* **128**: 145–153.
- Du Toit ED, May RM, Halliday IL, Schlaphoff T, Taljaard DG (1989). Paternity exclusion using 18 genetic systems in 2 124 cases in four South African population groups. *S Afr Med J* **75**: 103-105.
- Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evol Bioinf Online* **1**: 47-50.
- Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R, *et al.*, (1998). Jefferson fathered slave's last child. *Nature* **396**: 27-28.
- Gaulin SJC, Schlegel A (1980). Paternal confidence and paternal investment: a cross cultural test of a sociobiological hypothesis. *Ethol Sociobiol* **1**: 301-309.
- Geary DC (2006). Coevolution of paternal investment and cuckoldry in humans. In: Platek SM, Shackelford TK (eds) *Female infidelity and paternal uncertainty: evolutionary perspective on male anti-cuckoldry tactics*, Cambridge University Press: New York. pp 14-34.
- Giliomee H (2003). *The Afrikaners: biography of a people*. Tafelberg Publishers Limited: Cape Town.
- Giliomee H (2010). "Allowed such a state of freedom": Women and gender relations in the Afrikaner community before enfranchisement in 1930. *New Contree* **59**: 29-60.
- Giliomee H, Mbenga B (2007). *New history of South Africa*. Tafelberg: Cape Town.
- Goedbloed M, Vermeulen M, Fang RXN, Lembring M, Wollstein A, Ballantyne K *et al.*, (2009). Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFISTR (R) Yfiler (R) PCR amplification kit. *Int J Legal Med* **123**: 471-482.
- Goetz AT, Shackelford TD (2006). Sexual coercion and forced in-pair copulation as anti-cuckoldry tactics in humans. In: Platek SM, Shackelford TK (eds) *Female infidelity and paternal uncertainty: evolutionary perspective on male anti-cuckoldry tactics*, Cambridge University Press: New York. pp 82-99.
- Goetz AT, Shackelford TD (2009a). Sexual conflict in humans: evolutionary consequences of asymmetric parental investment and paternity uncertainty. *Anim Biol* **59**: 449-456.
- Goetz AT, Shackelford TD (2009b). Sexual coercion in intimate relationships: a comparative analysis of the effects of women's infidelity and men's dominance and control. *Arch Sex Behav* **38**: 226-234
- Goetz AT, Shackelford TK, Platek SM, Starratt VG, McKibbin WF (2007). Sperm competition in humans: implications for male sexual psychology, Physiology, Anatomy, and behavior. *Ann Rev Sex Res* **18**: 1-22.
- Gomendio M, Harcourt AH, Roldán ERS (1998). Sperm competition in mammals. In: Birkhead TR, Møller AP (eds) *Sperm competition and sexual selection*, Academic Press: London. pp 667-755.

- Gouws NB (1981). *Bevolkingstendense van Blankes in Suid-Afrika voor 1820*. PhD Thesis: Randse Afrikaanse Universiteit.
- Greeff JM (2007). Deconstructing Jaco: Genetic heritage of an Afrikaner. *Ann Hum Genet* **71**: 674-688.
- Greeff JM, Erasmus JC (2013). Appel Botha ~~Cornelitz~~: The abc of a three hundred year old divorce case. *Forensic Sci Int-Genet* **7**: 550-554.
- Greeff JM, Greeff FA, Greeff AS, Rinken L, Welgemoed DJ, Harris Y (2012). Low nonpaternity rate in an old Afrikaner family. *Evol Hum Behav* **33**: 268-273.
- Greene PJ (1978). Promiscuity, paternity and culture. *Amer Ethnol* **5**: 151-159.
- Groenewald G (2007). 'A mother makes no bastard': Family law, sexual relations and illegitimacy in Dutch colonial Cape Town, c. 1652-1795. *Afr Hist Rev* **39**: 58-90.
- Heese JA (1971). *Die Herkoms van die Afrikaner, 1657-1867*. A.A. Balkema: Kaapstad.
- Helgason A, Hrafnkelsson B, Gulcher JR, Ward R, Stefánsson K (2003). A populationwide coalescent analysis of Icelandic matrilineal and patrilineal genealogies: Evidence for a faster evolutionary rate of mtDNA lineages than Y chromosomes. *Am J Hum Genet* **72**: 1370-1388.
- Heyer E, Puymirat J, Dieltjes P, Bakker E, de Knijff P (1997). Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum Molec Genet* **6**: 799-803.
- Jobling MA (2001). In the name of the father: surnames and genetics. *Trends Genet* **17**: 353-357.
- Kayser M, Sajantila A (2001). Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic Sci Int* **118**: 116-121.
- Kayser M, Vermeulen M, Knoblauch H, Schuster H, Krawczak M, Roewer L (2007). Relating two deep-rooted pedigrees from Central Germany by high-resolution Y-STR haplotyping. *Forensic Sci Int-Genet* **1**: 125-128.
- Koehler N, Chisholm JS (2007). Early psychosocial stress predicts extra-pair copulations. *Evol Psych* **5**: 184-201.
- Kokko H, Jennions MD (2008). Parental investment, sexual selection and sex ratios. *J Evol Biol* **21**: 919-948.
- King TE, Jobling MA (2009a). Founders, drift, and infidelity: The relationship between Y chromosome diversity and patrilineal surnames. *Molec Biol Evol* **26**: 1093-1102.
- King TE, Jobling MA (2009b). What's in a name? Y chromosomes, surnames and the genetic genealogy revolution. *Trends Genet* **25**: 351-360.
- Larmuseau MHD, Vanoverbeke J, van Geystelen A, Defraene G, Vanderheyden N, Wenseleers KMT *et al.*, (2013). Low historical rates of cuckoldry in a western European human population traced by Y-chromosome and genealogical data. *Proc R Soc B* **280**: 20132400.
- Larmuseau MHD, Vanderheyden N, Van Geystelen A, van Oven M, de Knijff P, Decorte R (2014). Recent radiation within Y-chromosomal haplogroup R-M269 resulted in high Y-STR haplotype resemblance. *Ann Hum Genet* **78**: v92-103.
- Macintyre S, Sooman A (1991). Non-paternity and prenatal genetic screening. *Lancet* **338**: 869.
- McEvoy B, Bradley DG (2006). Y-chromosomes and the extent of patrilineal ancestry in Irish surnames. *Hum Genet* **119**: 212-219.
- Pagel M (1997). Desperately concealing father: a theory of parent-infant resemblance. *Anim Behav* **53**: 973-981.

- Parker GA (1982). Why are there so many tiny sperm? Sperm competition and the maintenance of two sexes. *J Theoret Biol* **96**: 281-294.
- Pollin TI, McBride DJ, Agarwala R, Schaffer AA, Shuldiner AR, Mitchell BD *et al.*, (2008). Investigations of the Y chromosome, male founder structure and YSTR mutation rates in the Old Order Amish. *Hum Hered* **65**: 91-104.
- Pound N, Shackelford TK, Goetz AT (2006). Sperm competition in humans. In: Shackelford TK, Pound N (eds) *Sperm competition in humans*, Springer: New York. pp 3-31.
- Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C *et al.*, (2010). Strong maternal Khoisan contribution to the South African Coloured population: a case of gender-biased admixture. *Am J Hum Genet* **86**: 611–620.
- R Development Core Team (2010). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing: Vienna.
- Rolf B, Keil W, Brinkmann B, Roewer L, Flimmers R (2001). Paternity testing using Y-str haplotypes: assigning a probability for paternity in cases of mutations. *Int J Legal Med* **115**: 12-15.
- Ridley M (2004). *Evolution*. Blackwell Science Ltd: Oxford.
- Sasse G, Müller H, Chakraborty R, Ott J (1994). Estimating the frequency of nonpaternity in Switzerland. *Hum Hered* **44**: 337-343.
- Scelza BA (2011). Female choice and extra-pair paternity in a traditional human population. *Biol Lett* **7**: 889-891.
- Shackelford TK, LeBlanc GJ (2001). Sperm competition in insects, birds, and humans: insights from a comparative evolutionary perspective. *Evol Cognit* **7**: 194-202.
- Shackelford TK, Pound N, Goetz AT (2005). Psychological and physiological adaptations to sperm competition in humans. *Rev Gen Psychol* **9**: 228-248.
- Shackelford TK, LeBlanc GJ, Weekes-Shackelford VA, Bleke-Rechek AL, Euler HA, Hoier S (2002). Psychological adaptation to human sperm competition. *Evol Hum Behav* **23**: 123-138.
- Simmons LW, Firman REC, Rhodes G, Peters M (2004). Human sperm competition: testis size, sperm production and rates of extrapair copulations. *Anim Behav* **68**: 297-302.
- Smith RL (1984). Human sperm competition. In: Smith RL (ed) *Sperm competition and the evolution of animal mating systems*, Academic Press, Inc: New York. pp 601-659.
- Soodyall H, Nebel A, Morar B, Jenkins T (2003). Genealogy and genes: tracing the founding fathers of Tristan da Cunha. *Eur J Hum Genet* **11**: 705-709.
- Starratt VG, Shackelford TK, Goetz AT, McKibbin WF (2007). Male mate retention behaviors vary with risk of partner infidelity and sperm competition. *Acta Psychol Sinica* **39**: 523-527.
- Strassmann BI, Kurapati NT, Hug BF, Burke EE, Gillespie BW, Karafet TM *et al.*, (2012). Religion as a means to assure paternity. *Proc Natl Acad Sci USA* **109**: 9781-9785.
- Sykes B, Irwen C (2000). Surnames and the Y chromosome. *Am J Hum Genet* **66**: 1417-1419.
- Trivers RL (1972). Parental investment and sexual selection. In: Campbell B (ed) *Sexual selection and the descent of man 1871-1971*, Aldine Publishing Company: Chicago. pp 136-179.
- Vermeulen M, Wollstein A, van der Graag K, Lao O, Xue Y, Wang Q *et al.* (2009). Improving global and regional resolution of male lineage differentiation by simple single-

- copy Y-chromosomal short tandem repeat polymorphisms. *Forensic Sci Inter-Genet* **3**: 205-213.
- Voracek M, Fisher M, Shackelford TK (2009). Sex differences in subjective estimates of non-paternity rates in Austria. *Arc Sex Behav* **38**: 652-656.
- Voracek M, Haubner T, Fisher ML (2008). Recent decline in nonpaternity rates: a cross-temporal meta-analysis. *Psychol Reports* **103**: 799-811.
- Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, *et al.*, (2001). The Eurasian Heartland: a continental perspective on Y-chromosome diversity. *Proc Nat Acad Sci USA* **98**: 10244-10249.
- Willuweit S, Roewer L (2007). Y chromosome haplotype reference database (YHRD): Update. *Forensic Sci Inter-Genet* **1**: 83-87.
- Wlodarski R, Manning J, Dunbar RIM (2015) Stay or stray? Evidence for alternative mating strategy phenotypes in both men and women. *Biol Let* **11**: 20140977.
- Wolf M, Musch J, Enczmann J, Fischer J (2012). Estimating the prevalence of nonpaternity in Germany. *Hum Nat* **23**: 208-217.

Supplementary Material for “Three hundred years of low non-paternity in a human population” by Jaco M. Greeff & J. Christoff Erasmus

Supplementary Table 1 Genealogically linked families genotyped in this study. The origin and year of arrival for each immigrant is given.

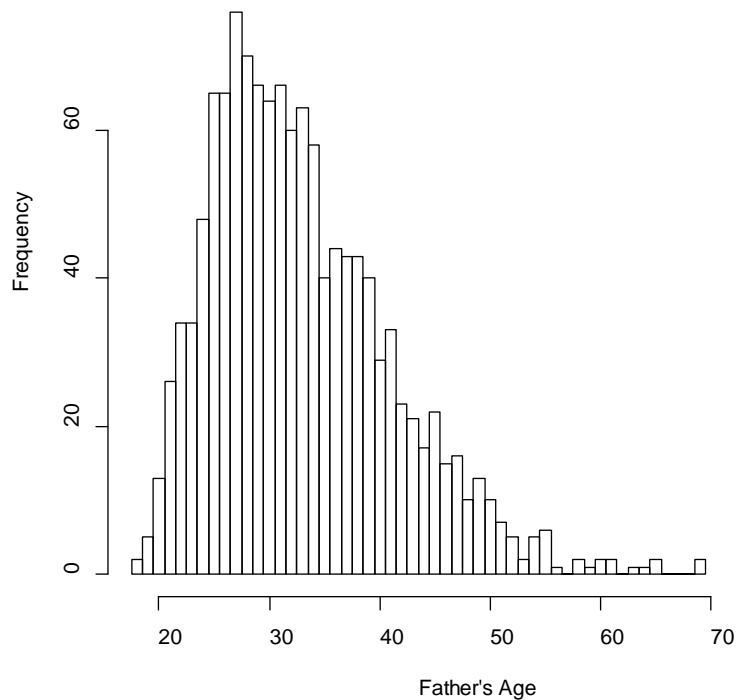
Origin	Surname	Arrival in Cape	Number of fertilizations	Total fertilizations
Dutch	Family 1	1679	44	
	Family 2	1693	54	
	Greeff, Cornelis J. ^a	1856 (born)	5	
	Family 3	1666	56	
	Family 4	1669	165	
	Family 5	before 1681	16	
	Family 6	1683 (born)	36	
	Family 7	1699	62	
	Family 8	1661	65	503
German	Family 9	1765	75	
	Botha, Friedrich ^b	1678	70	
	Family 10	before 1683	51	
	Greeff Matthias ^a	1680	103	
	Family 11	1713	62	
	Family 12	1765	14	
	Family 13	1663	24	
	Family 14	1688	105	504
French	Family 15	1698	39	
	Family 16	1710	41	
	Family 17	1688	29	
	Family 18	1718	76	
	Family 19	1671	25	210
Scandinavia	Family 20	1703	56	56
Total				1273

^a From Greeff *et al.* (2012)

^b From Greeff and Erasmus (2013). Due to well-known non-paternity in this family (Franken, 1926), we counted Maria Kickers’s first son as a non-paternity, but not her later sons as the married father seems to have consented to adultery and all the sons used their biological father’s surname (Greeff and Erasmus, 2013). In our total tally of conceptions we excluded the first generation of this family so that the denominator is 1273 rather than 1276.

Supplementary Table 2 Unique haplotypes found in this study.

Please see separate Excel file for this table. These data are also available on Dryad.



Supplementary Figure 1 A histogram of father's age at birth of their sons.

R code for calculating exclusion probabilities sampling back in time, as is done for coalescence.

This is to use coalescence back in time given a set of population sizes and number of immigrants for each time step. Note that there is an exclusion probability for each generation and that coalescence can happen in any of the previous time steps.

```
g1 <- 1000000 #number of iterations
trek <- 8 # generation where first two males are randomly drawn without replacement.
# Next follow the number of immigrants and the total population size in every interval, this
will be unique for each study
Immi<- c(68, 298, 317, 553, 758, 970, 2090, 0)
T <- c(68, 426, 770, 1789, 3798, 13008, 33493, 86232)
gen <- length(T)
c <- numeric(trek) # To remember in which generation two samples coalesce
```



```

d1 <- 0 #number of simulations counted

while (d1<g1) {#starting the cycle

#drawing the first two samples

e1 <- trek
n1<- sort(sample(1:T[e1],2,replace = FALSE)) #with replacement in the first sample only
if (n1[1] <= Immi[e1]) {e1 <- 0} #If either of the samples is a new immigrant it cannot
coalesce
else {
  e1 <- e1 - 1

#now drawing deeper samples

while (e1 > 0) {
  n1<- sort(sample(1:T[e1],2,replace = TRUE))
  if (n1[1] == n1[2]) {
    c[e1] <- c[e1] + 1
    e1 <- 0}
  else {
    if (n1[1] <= Immi[e1]){
      e1 <- 0}
    else {
      e1 <- e1 - 1}
    }
  }
}
d1 <- d1 + 1
}

# then the results of the simulation is given as follows:
# the exclusion probability:
1-sum(c)/d1
#The generation where coalescence occurred
c

```

References

Franken JLM (1926) Die Franse vlugteling. *Huisgenoot*, 16 July, 36-38.

Greeff JM, Erasmus JC (2013) Appel Botha ~~Cornelitz~~: The abc of a three hundred year old divorce case. *Forensic Sci Int-Genet* **7**: 550-554.

Greeff JM, Greeff FA, Greeff AS, Rinken L, Welgemoed DJ, Harris Y (2012) Low nonpaternity rate in an old Afrikaner family. *Evol Hum Behav* **33**: 268-273.