

# Deconstructing Jaco: Genetic Heritage of an Afrikaner

J. M. Greeff<sup>1,\*</sup>

<sup>1</sup>Department of Genetics, **University of Pretoria**, Pretoria 0002, South Africa

\*Correspondence author: Jaco M. Greeff, Fax: + 27 12 362 5327. E-mail:

jaco.greeff@up.ac.za

**[Figures and Tables at the bottom of the document]**

## Summary

It is often assumed that Afrikaners stem from a small number of Dutch immigrants. As a result they should be genetically homogeneous, show founder effects and be rather inbred. By disentangling my own South African pedigree, that is on average 12 generations deep, I try to quantify the genetic heritage of an Afrikaner. As much as 6% of my genes have been contributed by slaves from Africa, Madagascar and India, and a woman from China. This figure compares well to other genetic and genealogical estimates. Seventy three percent of my lineages coalesce into common founders, and I am related in excess of 10 times to 20 founder ancestors (30 times to Willem Schalk van der Merwe). Significant founder effects are thus possible. The overrepresentation of certain founder ancestors is in part explained by the fact that they had more children. This is remarkable given that they lived more than 300 years (or 12 generations) ago. DECONSTRUCT, a new program for pedigree analysis, identified 125 common ancestors in my pedigree. However, these common ancestors are so distant from myself, paths of between 16 and 25 steps in length, that my inbreeding coefficient is not unusually high ( $f \approx 0.0019$ ).

## Introduction

‘After three centuries of evolution the population structure of the Afrikaners is still far from stable, and there does not appear to be much prospect of its ever attaining uniformity. ... The numerous and often mutually contradictory genetic statements frequently made about them can consequently all be simultaneously true. The Afrikaner is a product of miscegenation, the last ‘pure European’, pathologically inbred and a manifestation of hybrid vigour, all at the same time.’ (Nurse et al. 1985)

Afrikaners are often considered a rather homogeneous, probably rather inbred, white population of Dutch ancestry. Yet, as the above quotation illustrates, there are uncertainties about the genetic composition of Afrikaners. Due to Afrikaners' high linkage disequilibrium, they are seen as a fruitful hunting ground for genes associated with disease (Hall et al. 2002). It is thus important that we have a clear appreciation of the Afrikaners' genetic heritage. In what follows I address the questions of racial admixture, nationalities, founder effects and inbreeding in the Afrikaner. I do so in a novel way: rather than taking a sample of modern Afrikaners and genotyping them, I start with one living Afrikaner and trace most of his South African ancestors. In this way I cast a net into his past and hope to get an impression of what the genetic heritage of a typical Afrikaner may be.

In lectures I often jokingly suggest that the Afrikaner population is a good example of inbreeding. The fact that Afrikaans children address any adults, related or not, as ‘oom’ and ‘tannie’, the Afrikaans for uncle and aunt, lends some credence to this suspicion. However, it was with mixed feelings that I, an Afrikaner, realized that the Afrikaner population has become a text book example of founder effects (Ridley, 2004). A founder effect refers to the phenomenon where the gene frequency of a new population is very different from that of its parent population. This happens when one or more of the immigrants have an unusual genotype and when the total number of immigrants are few. Although Ridley's suggestion that almost all Afrikaners stem from one ship-load of Dutch immigrants, who landed at the Cape of Good Hope in 1652, is most certainly incorrect he may well be correct about Afrikaners being a good example of founder effects. This simplistic one-ship-load suggestion is a common sentiment in many recent papers on Afrikaners. The unusually high incidence of a number of familial diseases

among Afrikaners suggests that there has been a significant founder effect (Botha & Beighton, 1983a, b; Nurse et al. 1985). However, de Villiers & Pama (1966c) list just over 4000 founders who emigrated to South Africa between 1652 and 1806. Even with such a large number of emigrants significant founder effects are possible, when emigration occurred over a long time, so that earlier arrivals contributed disproportionately more to the population. For instance, a number of founder effects in Saguenay-Lac Saint Jean (Quebec, Canada; Heyer & Tremblay, 1995; 1997) can be explained by cultural transmission of fitness (Austerlitz & Heyer, 1998; Heyer et al. 2005). Data collected by Heese (1971) show how this is possible in the case of Afrikaners (Table 1). For the 150 years recorded in Table 1, it is clear that there was a steady stream of emigration rather than a single event. Also notice how, over the years, more and more emigrants got married to locally born individuals. This is a typical situation where early emigrants can contribute disproportionately to the gene pool, leading to founder effects. Several diseases with an unusually high frequency in Afrikaners have been suggested to be the result of such founder effects: porphyria variegata (Dean, 1963), Beukes familial hip dysplasia (Cilliers & Beighton, 1990), familial hypercholesterolemia type I (Jenkins et al. 1980), Huntington's chorea (Hayden et al. 1980), Fanconi anaemia (Rosendorff et al. 1987), pseudoxanthoma elasticum (Torrington & Viljoen 1991), progressive familial heart block type I (Torrington et al. 1986), lipoid proteinosis (Heyl, 1970) and sclerosteosis (Beighton et al. 1977). Several studies have aimed to identify the actual founder ancestors who brought these disease-causing alleles to South Africa. Curiously, the founder Willem Schalk van der Merwe and his wife Elsje Cloete have been credited with introducing at least four diseases to the Afrikaner: Huntington's chorea (Hayden et al. 1980), pseudoxanthoma elasticum (Torrington & Viljoen, 1991), lipoid proteinosis (Heyl, 1970) and schizophrenia (Karayiorgou et al. 2004).

A founder effect normally goes hand in hand with elevated levels of inbreeding. Nurse *et al.* (1985) suggest that there may well be medically important levels of inbreeding in the Afrikaner population, stemming not only from the small population but also from a tendency to marry local people. The degree of inbreeding is normally calculated as the inbreeding coefficient (Wright, 1922),  $f$ , which is the chance that two alleles taken from a locus of an individual are identical by descent. In humans from Western societies close

kin unions (uncle-niece and first cousin) are uncommon (Bittles, 2001), and the inbreeding coefficient in Western populations tends to be very low, typically in the order of 0.001 (Bodmer & Cavalli-Sforza, 1976). In isolated populations of small size and originating from few founders, the inbreeding coefficient can be in excess of 0.01 (Mange, 1964; Bodmer & Cavalli-Sforza, 1976). In contrast, consanguineous unions are very common in west, central and south Asia, and north Africa (Bittles, 2001).

By comparing observed and expected heterozygosity, Lucassen (2005) found that the inbreeding coefficient of white South Africans is slightly negative ( $-0.001$ ). Since this estimate includes not just Afrikaners, it may not be entirely reflective of the Afrikaner population. However, if there is a significant substructure in the white population, then a Wahlund effect may in fact push this empirical estimate of Lucassen higher than it really is. This suggests that the current mating pattern is panmixia. To my knowledge there has not been extensive calculation of inbreeding coefficients for Afrikaners based on pedigrees. However, founder effects can result in a positive pedigree inbreeding coefficient even if the population is panmictic (Jacquard, 1975).

Given that genealogists could show that as much as 7% of Afrikaner genetic heritage is not of European descent (Heese, 1971), I find it curious that a system such as apartheid worked in South Africa. Seven percent is not a trivial amount, and is equivalent to having slightly more than a great-great-grandparent who was non-European. Since most of this non-European genetic heritage came into the Afrikaner population via female slaves, one would expect that as much as 14% of Afrikaner mitochondrial DNA is not even European. This female bias influx stems from the fact that emigrants were predominantly male, resulting in a male biased sex ratio of adults (Gouws, 1981).

Similarly, genetic studies also give support for this mixed racial ancestry. Working with a number of blood group gene frequencies, Botha & Pritchard (1972) estimated that between 6–7% admixture between western European and slaves from Africa and the East, and/or Khoikhoi, would be required to explain the allele frequencies. Nurse *et al.* (1985) listed a number of alleles typical to the Khoisan and Bantu-speaking peoples that are found in low frequencies in Afrikaners (ABO system:  $A^{\text{bantuu}}$ ; glucose-6-phosphate dehydrogenase:  $Gd^{\text{A}-}$  and  $Gd^{\text{A}}$ ; Rhesus:  $R^{\circ}$ ; Haemoglobin C).

Another argument concerning the heritage of the Afrikaner has concerned the relative contributions of Dutch, German and French immigrants. Although earlier authors suggested a mostly Dutch origin, more recent work has suggested almost equal contributions from these three groups (Heese, 1971). However, as Pama (de Villiers & Pama, 1966a) points out, the ever moving borderlines in the 1600s, and the regional distribution of customs and people in the 1600s make the current distinction somewhat artificial.

Heese's (1971) method needs to be explained. He recorded all the wedding dates, number of fertile children and origin of each immigrant. He divided the period from 1657 to 1837 into six 30 year periods. Since people who came to South Africa earlier contributed more to the nation, he multiplied each "blood unit" (fertile child) from the respective periods by 32, 16, 8, 4, 2 and 1. This approach clearly makes some mistakes, but given the numbers of people involved would probably give an answer fairly close to reality. This calculation, however, will work for Afrikaners as a whole, but for any individual it may vary largely from his estimate. This is because a more recent ancestor will contribute a larger proportion of a focal individual's DNA, whereas that recent ancestor contributes less to the population.

Recently, a number of studies on the human mating system and life history have made effective use of old church records (Helle et al. 2002, 2004; Volland & Beise, 2002; Cavalli-Sforza et al. 2004; Pettay et al. 2005). Afrikaners are in the unique position that their genealogies from the 1600s up to the early 1800s are very well constructed and recorded in reference books (de Villiers & Pama, 1966a, b, c; Heese & Lombard, 1986, 1989, 1992a, b; GISA, 1999, 2001, 2002a, b, 2003, 2004a, b, 2005, 2006). This makes the Afrikaner population ideal for studies of this nature.

I took advantage of these resources and recorded my own ancestral charts up until my ancestors immigrated to South Africa, and I tried to give some clarity to the questions raised above: founder effect, inbreeding coefficient, nationality composition and racial composition. These statistics will strictly only apply to my siblings and myself but it is of value to see how these calculations differ from those of Heese (1971). Deviations will reflect the influences of recent emigrants, but may also point to cultural inheritance of fitness (Heyer et al. 2005).

Very early on in this investigation, it became clear that certain founder ancestors seem to have contributed a disproportionate amount of DNA to me. Such an effect could simply be the result of chance drift, but as the population grew so fast (despite three smallpox epidemics it averaged 2.8% per year for the period 1735–1800 (Gouws, 1981)) this is unlikely. Two further alternative explanations are possible: it could result from fitness differences in these founders, as was found for the Saguenay population (Austerlitz & Heyer, 1998; Heyer et al. 2005), or it could stem from local mating groups that were established by specific founder ancestors, which was again illustrated for the Saguenay-Lac Saint Jean region (Lavoie et al. 2005). The data I accumulated allow me to test the fitness hypothesis. To do so sensibly, we need a brief digression to human life history theory.

One of the central tenets of life history theory is that there is a tradeoff between the quantity and quality of offspring (Lack, 1947; Smith & Fretwell, 1974). In humans, this tradeoff has been demonstrated inconsistently, with some studies supporting its existence (Strassmann & Gillespie, 2002; Hagen et al. 2006; Penn & Smith, 2007) and others not (Pennington & Harpending, 1988; Borgerhoff Mulder, 2000). Theoretically, for a specific amount of resources a mother will have an optimal number of offspring. If all mothers have similar resources we can expect an inverse U relationship between the fertility of mothers and their fitness, with mothers at the extremes producing too few or too many offspring. On the other hand, if mothers vary in the amount of resources that can be channelled to offspring, and each mother produces the optimal number of offspring, we expect a positive linear relationship between fertility and fitness. This suggests that the outcome depends on resource variation in different communities, and that without controlling for resources interpretation may be ambiguous.

In these studies fitness is measured as the number of children reaching the age of 5, or 10 years or the number of grandchildren. Here we have a unique opportunity to test this tradeoff, by comparing the fertility of founders with their genetic contribution to a specific Afrikaner living  $\pm 12$  generations later.

## Materials and Methods

### Ancestral Chart

I recorded my own ancestral chart (pedigree chart) for all my ancestors in South Africa. An ancestral chart is simply a list of all one's ancestors showing parent-offspring relationships along male and female lines. I did so until a specific individual immigrated to South Africa. I will refer to these immigrants as founder ancestors. In certain cases where immigrants were one another's sibs (Table S1, supporting material), I recorded this information to make calculation of the inbreeding coefficient more accurate. An ancestral chart, even one such as this with a defined starting point of immigration to South Africa, is never truly completed. New information can lead to changes and certain links may never be found. In my own case an eighth of my pedigree is still incomplete due to a great grandfather, J.L.M. van der Merwe, who was orphaned during the Anglo-Boer war and for whom I cannot find ancestral links. Nevertheless, I decided to continue with this work despite the incompleteness of the chart. I do not think that the basic findings would change much, and I tried to correct for this shortfall explicitly in the calculation of the inbreeding coefficient.

To complete this task, I used a number of books and reference works (Hoge, 1946; de Villiers & Pama, 1966a,b,c; Heese, 1971, 2005; Heese & Lombard 1986, 1989, 1992a,b; le Roux, 1988; GISA 1999, 2001, 2002a,b, 2003, 2004a,b, 2005, 2006) and two web resources: 'South Africa's Stamouers' (<http://www.stamouers.com/>) and 'The first Van Wijks at the Cape of Good Hope' ([http://www.ballfamilyrecords.co.uk/notes/VanWijk\\_intro.htm](http://www.ballfamilyrecords.co.uk/notes/VanWijk_intro.htm)). Mr. Henri Schoeman gave me the ancestors of A.C.C. Schoeman, and a document prepared by Mr. Hercules Malan allowed me to link my mother to known genealogies. In addition, interviews with my grandmother, aunt, uncle and parents filled in some gaps. The tombstones in two family graveyards, Greeff on the farm *Hazenjacht* and Maree on *Middelplaas*, both near De Rust, contained valuable links. Finally, for links I could not make using these sources I employed a professional genealogist, Mrs. Isabel Groesbeek. All pedigree data were typed into RootsMagic.

## **Composition**

To determine my genetic composition, in terms of the contributions of specific founder ancestors, I used the program Deconstruct (see Appendix) which takes Gedcom files as an input. Information on race and country of origin of the founder ancestors were obtained from the same resources that were used to draw up the ancestral chart. These could then be added together, according to race or nationality.

## **Inbreeding Coefficient and Founder Effect**

Studies using pedigrees to calculate inbreeding in humans normally only consider up to the third (Cavalli-Sforza et al. 2004) or fourth cousin (Mange, 1964) relationships. This is because more distant relationships add very little to the inbreeding coefficient ( $<1/1024$ ). In this case I was mainly interested in the effects of longer paths between my parents, in the order of 12 generations ago. These would each add  $\frac{1}{4}^{23}$  to the inbreeding coefficient, but there may potentially be many such paths. The program Deconstruct (see Appendix) was used to calculate my inbreeding coefficient. Due to the absence of my one great grandfather, inbreeding accumulation curves were estimated with the program Deconstruct. Thirty random sequences of great grandparents were investigated, each time running  $10^6$  simulations. A regression line fitted to the inbreeding coefficient versus the number of great-grandparents considered can then be used to extrapolate a value for the inbreeding coefficient if all my great grandparents are considered.

Fast population growth can reduce the amount of inbreeding in a small founded population. A founder effect can thus be hard to illustrate with reference to the inbreeding coefficient. It is easier to look at ancestor loss to get some impression of a founder effect. One expects  $2^x$  ancestors  $x$  generations ago. For example, I should have 2048 ancestors 11 generations ago and double that 12 generations ago. Clearly there were fewer people in the Cape in the 1600s, so some ancestors are not unique and I am related to them several times. This ancestor loss was calculated using Deconstruct.

## **Fitness**

For the period 1657–1687 I took all the couples listed in Heese (1971), and recorded their number of children from de Villiers & Pama (1966a,b,c) under the name of the father. In

one case, where a male was married twice, the offspring from both marriages were counted. Heese (1971) only recorded children that survived, remained in the Cape, and had children. Since the point is to see if families with more children have in fact a lower fitness, due to the death of some children, it would be incorrect to use his values. I also recorded the marriage date or, if this was not given or they were not married, I took it as being two years prior to the birth of their first offspring. The output from Deconstruct was then compared to this list of families; specifically, the number of times I am related to each person, and my relatedness to each person was recorded. These two values were thus taken as proxies for fitness. I then fitted a generalized linear model with Poisson errors to the data, with times related as the dependant variable and year of marriage, number of offspring and number of offspring squared as the independent variables. The marriage date was included because the time window considered, 30 years, is substantial enough to cause a difference between earlier and later people. Specifically, one expects to be related more times to earlier arrivals, but to be less related for any one link. The square of the number of offspring was fitted to allow a u-shape to be retrieved. A linear model with relatedness as the dependent and the same independent variables was also fitted. Non-significant terms were removed until the minimum adequate model remained (Crawley, 2005). These analyses were done in R (R Development Core Team, 2005).

## **DNA**

Studies based on ancestry will be incorrect when the father had been cuckolded, or if children were adopted but this fact has not been recorded. In a number of cases female founder ancestors were simply denoted as van die Kaap, meaning from the Cape, which is understood as slaves born at the Cape. Thus, to confirm a few of the proposed lineages and to determine from where these 'van die Kaap' females came from, a number of mitochondrial DNA (mtDNA) and Y chromosomes were typed. My own Y chromosome, stemming from Matthias Greeff from Magdeburg (Germany), my own mtDNA, donated by Claudine Eloy (Cloy) from Bordighera (Liguria Italy), and my aunt's mtDNA, donated by Maria Bastiaans van die Kaap, were typed. This was done by the Human Genomic Diversity and Disease Research Unit, sequencing HVRI and HVRII of the mtDNA and scoring the bi-allelic markers as well as 7 STRs on the Y-chromosome.

## Results

The completed pedigree contained 926 individuals. The longest and shortest completed lineages were 14 and 5 generations long respectively, excluding myself. In addition to the ancestors of my great-grand father, J.L.M. van der Merwe, six other ancestors could not be followed to the point of immigration to South Africa (Table S2). Together they account for 17.2% of my genes. The pedigree chart can be found in the Supplementary Material. Due to space constraints I will only list a few ancestors here, but the complete list can be found in the Supplementary Material (Table S3).

### Composition

For this pedigree 299 founder ancestors contributed to my genes. The accumulation of founder ancestors as additional great-grandparents were sequentially added suggests that there remain only 7 more founder ancestors to be discovered (Figure 1a).

Theoretically, a great-grandfather should contribute 256 ancestors 11 generations separated from me. This discrepancy, 7 versus 256, stems from the fact that many ancestors are related to me via several lineages. Although I am related to most of my ancestors once only (Figure 2a), to many I am related several times (Table 2). Most notable is Willem Schalk van der Merwe and his wife's parents, to whom I am related 30 times. As a result I have a fairly high relatedness to some ancestors that are on average 12 generations removed from myself (Table 3). However, my highest relatedness is to a few immigrants that came to South Africa more recently (Table 3). Even so, to most founder ancestors I am related by less than 0.25% (Figure 2b).

The contributions from different nationalities and European versus non-European founder ancestors are given in Table 4. I have more French and less German ancestors than Heese (1971) calculated to be average for Afrikaners. About 6% of my genes were contributed by non-European founder ancestors. These ancestors are listed in Table 5.

Excluding incomplete lineages, and counting van die Kaap individuals from when they were recorded, the average lineage is 10.8 generations long, i.e. I am the 11.8<sup>th</sup> generation. Similarly my average gene has spent 9.6 generations in South Africa before it

was passed on to me. Figure 2c gives a frequency distribution of the number of generations between me and my founder ancestors.

### **Founder Effect and Inbreeding Coefficient**

I am related to my 299 founder ancestors 1101 times. This discrepancy results from certain ancestors being hit multiple times (as seen above). There is thus a large potential for founder effects caused by certain founder ancestors' alleles, such as van der Merwe for instance.

Considering only 7 of my great grandparents,  $5 \times 10^9$  simulations gave my inbreeding coefficient as 0.001511. Looking at how my inbreeding coefficient increases with the random addition of great grandparents my predicted inbreeding coefficient if all 8 great grandparents were to be considered is  $\pm 0.0019$  (Figure 1b).

During simulations Deconstruct recorded the path lengths and identity of the common ancestor for 65224 simulations where a common ancestor was encountered. One hundred and 25 common ancestors were identified, with the shortest path equal to 16 steps and the longest path equal to 25 steps. This is equal to sixth cousins once removed and 11<sup>th</sup> cousins, respectively. Generation overlap is common, with 53 common ancestors having two path lengths, 17 three, 14 four and one five. The bimodal distribution of path lengths is given in Figure 2d. Twenty-six of the common ancestors account for 62% of all the inbreeding. These 26 common ancestors are, not surprisingly, the individuals to whom I am related many times and their offspring.

### **Fitness**

I am related to only 33 of the 63 families listed by Heese (1971) who were married between 1657 and 1687. The number of offspring of Visser and Willem van Wijk was updated with newer findings, reported on the web resources consulted. According to Pama the average number of offspring was 5.94 with a standard deviation of 3.04. The analysis with number of times related was overdispersed so a quasi-Poisson model was specified. The relatedness had to be square root transformed to improve the fit to model assumptions of the linear model. There was no support for an inverse U-shaped relationship between offspring number and the two fitness proxies (Table 6; Figure 3).

The number of times related, as well as relatedness, increased as the number of offspring of the founder ancestors increased (Table 6; Figure 3). In the case of number of times related, marriages that occurred earlier were more likely to result in a link with the ancestor (Table 6).

## **DNA**

My mtDNA fell into group W and my Y chromosome fell into haplogroup R; my father's mtDNA fell into group M.

## **Discussion**

Most approaches to human population genetics start by sampling a population. Here I have taken the opposite approach, and used one individual living in the present to sample more than a thousand entangled lineages running back into the past. In this way I have tried to elucidate information both about the founder ancestors of the Afrikaners, as well as present day Afrikaners.

Even though my pedigree is 17% incomplete, the accumulation curves suggest that as few as 7 more founder ancestors remain to be discovered. This is due to the substantial ancestor loss to which I will return below. It does mean that what I have calculated here with regards to my composition is unlikely to change substantially as additional lineages are completed. The fact that my average lineage, including myself, is almost 12 generations long and that the average gene in me is now spending almost its 11th generation on this continent show that, although there are some founder ancestors that immigrated here more recently, most contributions stem from the early immigrants. One may thus expect a fairly close agreement between my composition and the average Afrikaner as calculated by Heese (1971).

Considering the nationalities of contributors I contain 12% more French genetic heritage than Heese's calculations (Table 3). This increase is balanced by an 8% loss from German, a 3% loss from other European, and a 1% loss from non-European genetic heritage. The fact that most non-European genetic heritage came into the Afrikaner population via German immigrants (Heese, 2005) probably explains the drop in non-European genetic heritage. However, my high relatedness to the more recent immigrant,

Rosalyn van Macao, who contributed 25% of my non-European genetic heritage, 'reconciled' this mismatch between my data and that of Heese's (1971).

It is not clear if my higher estimate of French contribution is because of a systematic mistake in Heese's (1970) estimate, or if it is because of a quirkiness in my own ancestry. It seemed to be the case that when a lineage hit the French Huguenots it stayed in this group. It will be interesting to compare the degree of inbreeding of the early generations of Huguenots to the other early immigrants. In the light of the calculations of Heyer et al. (2005) there is an interesting possibility that the cultural inheritance of fitness may have led to a systematic bias in Afrikaners, since Huguenots tended to be more educated and trained than German emigrants who tended to be soldiers. We are currently investigating this hypothesis.

It is unclear how the 2% contribution by women known only as van die Kaap should be allocated. At least one of them, Maria Bastiaans van die Kaap, carried an M mtDNA haplotype, which suggests an Asian and possibly Indian ancestry (Maca-Meyer et al. 2001). Although more slaves came to the Cape from Madagascar than from Guinea, Guinea contributed almost 5 times more to my gene pool than Madagascar. This is through the fertile contributions of a woman known as Lijsbeth Sanders van die Kaap. From her social associations it is believed that she was from Guinea (Hattingh, 1980). Of the slaves, however, a large number of women from India contributed to my genes. My estimate of 6% non-European genetic contribution is in close agreement with genealogical (Heese, 1971) and genetic (Botha & Pritchard, 1972) estimates. I am not aware that this estimate has been validated for any other Afrikaner individual, but it will be interesting if this can be confirmed for more Afrikaners. Presently, most white and black South Africans are equally incredulous at the prospect that Afrikaners have such a rich genetic heritage. Hopefully, with time all South Africans will celebrate the fact that Afrikaners are, and continue to be, a proudly south African concoction.

The unusually large linkage disequilibrium in Afrikaners (Hall et al. 2002) is thus explained by this heterogeneous starting population, as well as the relatively few generations (12) since the origins of the Afrikaner.

Of 1101 lineage tips only 299 are unique founder ancestors. According to Pearl's (1917) inbreeding estimator, that measures the fraction of lineages that have coalesced, 73% of

my lineages coalesced within 12 generations. Pattison (2004) estimated that the same degree of coalescence would only have occurred between 1450 and 1600 for Britain, India, Japan, Europe and China. This suggests that the Afrikaner population is indeed small and suffers from a founder effect. However, Pattison (2004) assumed that the entire population of each of these areas mated at random, a very unlikely scenario, and this will certainly inflate these coalescence times.

With only 299 founder ancestors, and with some ancestors being so dominant in their contributions, it is very likely that founder effects could have played an important role. Two points need to be mentioned. First, if any of the people in Table 2 carried a disease allele it will now occur at a high frequency in Afrikaners. Second, Table 2 is a list of the usual suspects for introducing diseases, but it is not clear if this is because they were in fact fitter (see below) or whether they actually carried the diseases. Willem Schalk van der Merwe will probably be a common ancestor for any two Afrikaners and will thus always be a possible candidate. Further support for this cautionary message comes from the fact that my father and mother share 125 common ancestors. In other words, if they both carried the same familial disease, 125 people could have been the possible donor. One should thus be cautious to identify the actual founder individuals before complete pedigrees have been obtained.

In support of the fitness/carrier debate I can list the following examples where I am related to proposed carriers, often several times: first, Willem Schalk van der Merwe and/or his wife allegedly brought Huntington's chorea (Hayden et al. 1980), pseudoxanthoma elasticum (Torrington & Viljoen, 1991), lipoid proteinosis (Heyl, 1970) and schizophrenia (Karayiorgou et al. 2004) to South Africa, but I am related to him at least 30 times. Interestingly, Hayden (1980) argued that all the Huntington's chorea lines ran through Sophia van der Merwe, a daughter of Willem Schalk. Although I am related to Willem Schalk 30 times, this is through 4 of his 10 other children who left descendants in South Africa but not through Sophia. This suggest that this link may well be less likely than the regular van der Merwe link, and that Hayden et al. (1980) were correct to argue that this could have been due to a new mutation in Sophia.

Second, Ignasius Ferreira may have brought progressive familial heart block to South Africa (Torrington et al. 1986), but I am related to Ferreira twice. However, Ferreira's

wife was a Terblanche who on her mother's side was a le Febre. To these surnames I am related 17 times. Again, it becomes pretty likely that anyone may be linked to these founder ancestors. Similarly to Huntington's chorea, all the lineages run to Ferreira via one of his sons, Thomas Ignatius, suggesting that this may have been a new mutation too (van der Merwe et al. 1994).

Third, Gerrit Jansz van Deventer or his wife Arriaenje Jacobs Adriaanse apparently carried porphyria variegata (Dean, 1963). To van Deventer and his wife I am related 7 times with a further one link to Arriaenje's sister, Willemyntjie Ariens de Wit.

Fourth, there are four possible families, van der Merwe, Burger, Visser and Smit, who have been suggested as the possible origin of psedoxanthoma elasticum (Torrington & Viljoen, 1991). I am, respectively, related to these families 30, 12, 17 and two times.

It is thus clear that the usual suspects (Table 2) will be encountered more often if geneticists do not work with completed pedigrees, and are thus very likely to be incorrectly identified as the disease carriers. Researchers doing similar work on the Sagueneay population of Canada have also found a number of founders that contributed disproportionately to the population (Heyer & Tremblay, 1995; Heyer et al. 1997; Heyer, 1999). They corrected for this phenomenon by including an equivalent control group of healthy people (Heyer & Tremblay, 1995; Heyer et al. 1997; Vézina et al. 1999, 2005). Given the amount of ancestor loss it comes as a bit of a surprise that my inbreeding coefficient is only 0.0019. It is about twice as high as estimates for other Western countries, but these estimates are normally based on ancestries that are only three generations deep. All my inbreeding comes from much deeper common ancestors (Figure 2d). Given the small magnitude of the inbreeding coefficients calculated for Western societies, a mistake of 0.0019 is not trivial. Even though the founding of the Cape population seems very unlike stable communities in Europe, many communities in the 1600s were very stagnant, and similar founder effects could be important. This low level of inbreeding given the high ancestor loss is probably explained by the rapid expansion of the population (Gouws, 1981; Halliburton, 2004).

If generations did not overlap, one would expect path lengths to show a peak at every second path length because each generation adds two parents. The fact that it does not (Figure 2d) is the result of generation overlap. This can also be seen from the fact that so

many common ancestors were on paths of varying length, up to 5 steps in difference. Such generation overlaps can be expected for populations like these, where girls got married at a very young age and sometimes had more than ten children.

The fact that founder ancestors who had more children are more related to me, and related to me more times, vindicates our use of number of offspring as a measure of fitness, as originally proposed by Darwin (1859). One may have expected that the fitness values of the subsequent 12+ generations would have destroyed this signal. However, the cultural transmission of fitness (Heyer et al. 2005) would make this observation more likely. The inverse U relationship expected under a quantity-quality tradeoff was not found. This suggests that the quantity-quality tradeoff may not have existed in this population. The fact that the Afrikaner population grew almost 5 times faster per year than the average Scandinavian population at the same time (Gouws, 1981) implies that competition for resources may not have been limiting. However, we cannot exclude the possibility that larger families simply had more resources.

The DNA typing confirmed and clarified places of origin of certain founders. W is a relatively rare haplotype that is most common in Western Eurasia, reaching frequencies of 17% in the Sindhi population of South Eastern Pakistan (Quintana-Murci et al. 2004). However, W does occur in southern Europe (Torroni et al. 1996; Maca-Meyer et al. 2001) and is not at odds with Claudine Eloy from Bordighera being the carrier. The M haplotype of my father's mtDNA was donated by a slave born at the Cape. We can now speculate that her mother must have come from Asia, possibly India (Maca-Meyer et al. 2001), rather than from Africa. My Y chromosome is a common European genotype and could have come from Germany (Wells et al. 2001).

The approach followed in this study offers unique advantages, in that it links history to its genetic consequences. Although my ancestry sampled more than 1000 lineages, i.e. was potentially all inclusive, the high degree of ancestor loss may suggest that these conclusions are not generally true for Afrikaners. Of the potential 63 couples married before 1687 in the Cape I am related to only 33. This skew is in part explained by the fitness differences between the founders (Figures 3a & b), but these models were overdispersed suggesting that other factors are also important. I suspect that the extra noise could be explained by the fact that marriage partners were normally obtained from

the local population, so that place of origin can tie a number of families together. To answer these questions similar analyses needs to be conducted on more Afrikaners and geography needs to be incorporated.

## References

- Austerlitz, F. & Heyer, E. (1998) Social transmission of reproductive behavior increases frequency of inherited disorders in a young-expanding population. *Proc Natl Acad Sci USA* **95**, 15140–15144.
- Beighton, P., Davidson, J., Durr, L. & Hamersma, H. (1977) Sclerotosis - an autosomal recessive disorder. *Clin Genet* **11**, 1–7.
- Bittles, A. H. (2001) Consanguinity and its relevance to clinical genetics. *Clin Genet* **60**, 89–98.
- Bodmer, W. F. & Cavalli-Sforza, L. L. (1976) *Genetics, Evolution and Man*. San Francisco: W. H. Freeman and Company.
- Borgerhoff Mulder, M. (2000) Optimizing offspring: the quantity-quality tradeoff in agropastoral Kipsigis. *Evol Hum Behav* **21**, 391–410.
- Botha, M. C. & Beighton, P. (1983a) Inherited disorders in the Afrikaner population of southern Africa. Part I. Historical and demographic background, cardiovascular, neurological, metabolic and intestinal conditions. *S Afr Med J* **64**, 609–612.
- Botha, M. C. & Beighton, P. (1983b) Inherited disorders in the Afrikaner population of southern Africa. Part II. Skeletal, dermal and heamatological conditions; the Afrikaners of Gamkaskloof; demographic considerations. *S Afr Med J* **64**, 664–667.

- Botha, M. C. & Pritchard, J. (1972) Blood group gene frequencies: an indication of the genetic constitution of population samples in Cape Town. *S Afr Med J* **46**, S1–27.
- Boyce, A. J. (1983) Computation of inbreeding and kinship coefficients on extended pedigrees. *J Hered* **74**, 400–404.
- Cavalli-Sforza, L. L., Moroni, A. & Zei, G. (2004) *Consanguinity, Inbreeding, and Genetic Drift in Italy*. Princeton: Princeton University Press.
- Cilliers, B. (1976) A generalisation of an approximate method to calculate inbreeding coefficients. *S Afr J Anim Sci* **6**, 191–193.
- Cilliers, H. J. & Beighton, P. (1990) Beukes familial hip dysplasia: an autosomal dominant entity. *Am J Med Genet* **36**, 386–390.
- Cook, R. D. & Hartl, D. L. (1976) Estimation of inbreeding by random walks in pedigrees. *Theor Appl Genet* **48**, 171–178.
- Crawley, M. J. (2005) *Statistics: an introduction using R*. Chichester: John Wiley and Sons Ltd.
- Darwin, C. (1859) *The Origin of Species by means of Natural Selection*. London: John Murray.
- de Villiers, C. C. & Pama, C. (1966a) *Genealogies of Old South African Families*. Cape Town: A.A. Balkema.
- de Villiers, C. C. & Pama, C. (1966b) *Genealogies of Old South African Families*. Cape Town: A.A. Balkema.

•

de Villiers, C. C. & Pama, C. (1966c) *Genealogies of Old South African Families*.  
Cape Town: A.A. Balkema.

•

Dean, G. (1963) *The Porphyrias: a story of inheritance and environment*.  
Bath: The Pitman Press.

•

GISA (1999) *South African Genealogies 5, L-M*.  
Stellenbosch: Genealogical Institute of South Africa.

•

GISA (2001) *South African Genealogies 6, N*.  
Stellenbosch: Genealogical Institute of South Africa.

•

GISA (2002a) *South African Genealogies 7, O-Ph*.  
Stellenbosch: Genealogical Institute of South Africa.

•

GISA (2002b) *South African Genealogies 8, Pi-Q*.  
Stellenbosch: Genealogical Institute of South Africa.

•

GISA (2003) *South African Genealogies 9, Ra-Ron*.  
Stellenbosch: Genealogical Institute of South Africa.

•

GISA (2004a) *South African Genealogies 10, Roo-Schn*.  
Stellenbosch: Genealogical Institute of South Africa.

•

GISA (2004b) *South African Genealogies 11, Schno-Sny*.  
Stellenbosch: Genealogical Institute of South Africa.

•

GISA (2005) *South African Genealogies 12, So-Sz*.  
Stellenbosch: Genealogical Institute of South Africa.

•

GISA (2006) *South African Genealogies 13, T-U*.  
Stellenbosch: Genealogical Institute of South Africa.

•

Gouws, N. B. (1981) *Bevolkingstendense van Blankes in Suid-Afrika voor 1820*.  
Johannesburg: Randse Afrikaanse Universiteit.

•

Hagen, E. H., Barrett, H. C. & Price, M. E. (2006) Do human parents face a quantity-quality tradeoff?: Evidence from a Shuar community. *Am J Phys Anthr* **130**, 405–418.

•

Hall, D., Wijsman, E. M., Roos, J. L., Gogos, J. A. & Karayiorgou, M. (2002) Extended intermarker linkage disequilibrium in the Afrikaners. *Genome Res* **12**, 956–961.

•

Halliburton, R. (2004) *Introduction to Population Genetics*.  
Upper Saddle River: Pearson Prentice Hall.

•

Hattingh, J. L. (1980) Die Blanke nageslag van Louis van Bengale en Lijsbth van die Kaap. *Kronos* **3**, 5–51.

•

Hayden, M. R., Hopkins, H. C., MaCrae, M. & Beighton, P. H. (1980) The origin of Huntington's Chorea in the Afrikaner population of South Africa. *S Afr Med J* **58**, 197–200.

•

Heese, H. F. (2005) *Groep Sonder Grense*.  
Pretoria: Protea Boekhuis.

•

Heese, J. A. (1971) *Die Herkoms van die Afrikaner, 1657–1867*.  
Kaapstad: A.A. Balkema.

•

Heese, J. A. & Lombard, R. T. J. (1986) *South African Genealogies 1, A–C*.  
Pretoria: Human Sciences Research Council.

- Heese, J. A. & Lombard, R. T. J. (1989) *South African Genealogies 2, D–G*.  
Pretoria: Human Sciences Research Council.
- Heese, J. A. & Lombard, R. T. J. (1992a) *South African Genealogies 3, H–I*.  
Pretoria: Human Sciences Research Council.
- Heese, J. A. & Lombard, R. T. J. (1992b) *South African Genealogies 4, J–K*.  
Pretoria: Human Sciences Research Council.
- Helle, S., Käär, P. & Jokela, J. (2002) Human longevity and early reproduction in pre-industrial Sami populations. *J Evol Biol* **15**, 803–807.
- Helle, S., Lummaa, V. & Jokela, J. (2004) Selection for increased brood size in historical human populations. *Evol* **58**, 430–436.
- Heyer, E. (1999) One founder/one gene hypothesis in a new expanding population: Saguenay (Quebec, Canada). *Hum Biol* **71**, 99–109.
- Heyer, E., Sibert, A. & Austerlitz, F. (2005) Cultural transmission of fitness: genes take the fast lane. *Trends Genet* **21**, 234–239.
- Heyer, E. & Tremblay, M. (1995) Variability of the genetic contribution of Quebec population founders associated to some deleterious genes. *Am J Hum Genet* **56**, 970–978.
- Heyer, E., Tremblay, M. & Desjardins, B. (1997) Seventeenth-century European origins of hereditary diseases in the Saguenay population (Quebec, Canada). *Hum Biol* **69**, 209–225.
- Heyl, T. (1970) Genealogical study of lipoid proteinosis in South Africa. *Brit J Derm* **83**, 338–340.

•

Hoge, J. (1946) Personalia of Germans at the Cape, 1652–1806. *Argiefjaarboek vir Suid-Afrikaanse Geskiedenis (9de jaargang)*.

Kaapstad: Cape Times.

•

Jacquard, A. (1975) Inbreeding: one word, several meanings. *Theor Pop Biol* **7**, 338–363.

•

Jenkins, T., Nicholls, E., Gordon, E., Mendelsohn, D., Seftel, H. & Andrew, M. J. A. (1980) Familial hypercholesterolemia - a common genetic disorder in the Afrikaans population. *S Afr Med J* **57**, 943–947.

•

Karayiorgou, M., Torrington, M., Abecasis, G. R., Pretorius, H., Robertson, B., Kaliski, S., Lay, S., Sobin, C., Moller, N., Lundy, S. L., Blundell, M. L., Gogos, J. A. & Roos, J. L. (2004) Phenotypic characterization and genealogical tracing in an Afrikaner schizophrenia database. *Am J Med Genet B-Neuropsychiatric Genetics* **124B**, 20–28.

•

Lack, D. (1947) The significance of clutch size. *Ibis* **89**, 302–352.

•

Lavoie, E.-M., Tremblay, M., Houde, L. & Vézina, H. (2005) Demogenetic study of three populations within a region with strong founder effects. *Community Genet* **8**, 152–160.

•

le Roux, J. G. (1988) *Hugenotebloed in ons are*. Pretoria: Raad vir Geesteswetenskaplike Navorsing.

•

Lucassen, A. (2005) *Population genetics of STR loci in four population groups in South Africa*. Pretoria: Tshwane University of Technology.

•

Maca-Meyer, N., González, A. M., Larruga, J. M., Flores, C. & Cabrera, V. M. (2001) Major genomic mitochondrial lineages delineate early human expansions. *BMC Genetics* **2**, 13.

- 
- Mange, A. P. (1964) Growth and inbreeding of a human isolate. *Hum Biol* **36**, 104–133.
- 
- Nurse, G. T., Weiner, J. S. & Jenkins, T. (1985) *The Peoples of Southern Africa and their Affinities*. Oxford: Oxford University Press.
- 
- Pattison, J. E. (2004) A comparison of inbreeding rates in India, Japan, Europe and China. *Homo-J Comp Hum Biol* **55**, 113–128.
- 
- Pearl, R. (1917) Studies on inbreeding. VII. - Some further considerations regarding the measurement and numerical expression of degrees of kinship. *Am Nat* **51**, 545–559.
- 
- Penn, D. J., & Smith, K. R. (2007) Differential fitness costs of reproduction between the sexes. *Proc Nat Acad Sci USA* **104**, 553–558.
- 
- Pennington, R. & Harpending, H. (1988) Fitness and fertility among Kalahari !Kung. *Am J Phys Anthr* **77**, 303–319.
- 
- Pettay, J. E., Kruuk, L. E. B., Jokela, J. & Lummaa, V. (2005) Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *Proc Nat Acad Sci USA* **102**, 2838–2843.
- 
- Quintana-Murci, L., Chaix, R., Wells, R. S., Behar, D. M., Sayar, H., Scozzari, R., Rengo, C., Al-Zahery, N., Semino, O., Santachiara-Benerecetti, A. S., Coppa, A., Ayub, Q., Mohyuddin, A., Tyler-Smith, C., Mehdi, S. Q., Torroni, A. & McElreavey, K. (2004) Where west meets east: The complex mtDNA landscape of the southwest and Central Asian corridor. *Am J Hum Genet* **74**, 827–845.
- 
- R Development Core Team (2005). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.

•

Ridley, M. (2004) *Evolution*. Oxford: Blackwell Science Ltd.

•

Rosendorff, J., Bernstein, R., Macdougall, L. & Jenkins, T. (1987) Fanconi Anemia: another disease of unusually high prevalence in the Afrikaans population of South Africa. *Am J Med Genet* **27**, 793–797.

•

Smith, C. C. & Fretwell, S. D. (1974) The optimal balance between size and number of offspring. *Am Nat* **108**, 499–506.

•

Strassmann, B. I. & Gillespie, B. (2002) Life-history theory, fertility and reproductive success in humans. *Proc R Soc Lond B* **269**, 553–562.

•

Torrington, M. & Viljoen, D. L. (1991) Founder effect in 20 Afrikaner kindreds with pseudoxanthoma elasticum. *S Afr Med J* **79**, 7–11.

•

Torrington, M., Weymar, H. W., van der Merwe, P.-L. & Brink, A. J. (1986) Progressive familial heart block, part 1. Extent of the disease. *S Afr Med J* **70**, 353–355.

•

Torrioni, A., Huoponen, K., Francalacci, P., Petrozzi, M., Morelli, L., Scozzari, R., Obinu, D., Savontaus, M. L. & Wallace, D. C. (1996) Classification of European mtDNAs from an analysis three European populations. *Genet* **144**, 1835–1850.

•

van der Merwe, P.-L., Weymar, H. W. & Kalis, N. N. (1994) The extent of progressive familial heart block type 1 - a new perspective. *Card J S Afr* **5**, 125–126.

•

Vézina, H., Durocher, F., Dumont, M., Houde, L., Szabo, C., Tranchant, M., Chiquette, J., Plante, M., Laframboise, R., Lépine, J., BCLC Haplotype Group, Nevanlinna, H., Stoppa-Lyonnet, D., Goldgar, D., Bridge, P., INHERIT BRCAs & Simard, J. (2005) Molecular and genealogical characterization of the R1443X *BRCA1* mutation in high-risk French-Canadian breast/ovarian cancer families. *Hum Genet* **117**, 119–132.

•

Vézina, H., Heyer, E., Fortier, I., Quéllette, G., Robitaille, Y. & Gauvreau, D. (1999) A genealogical study of Alzheimer disease in the Saguenay region of Quebec. *Genet Epidemiol* **16**, 412–425.

•

Voland, E. & Beise, J. (2002) Opposite effects of maternal and paternal grandmothers on infant survival in historical Krummhörn. *Behav Ecol Sociobiol* **52**, 435–443.

•

Wells, R. S., Yuldasheva, N., Ruzibakiev, R., Underhill, P. A., Evseeva, I., Blue-Smith, J., Jin, L., Su, B., Pitchappan, R., Shanmugalakshmi, S., Balakrishnan, K., Read, M., Pearson, N. M., Zerjal, T., Webster, M. T., Zholoshvili, I., Jamarjashvili, E., Gambarov, S., Nikbin, B., Dostiev, A., Aknazarov, O., Zalloua, P., Tsoy, I., Kitaev, M., Mirrakhimov, M., Chariev, A. & Bodmer, W. F. (2001) The Eurasian Heartland: A continental perspective on Y-chromosome diversity. *Proc Nat Acad Sci USA* **98**, 10244–10249.

•

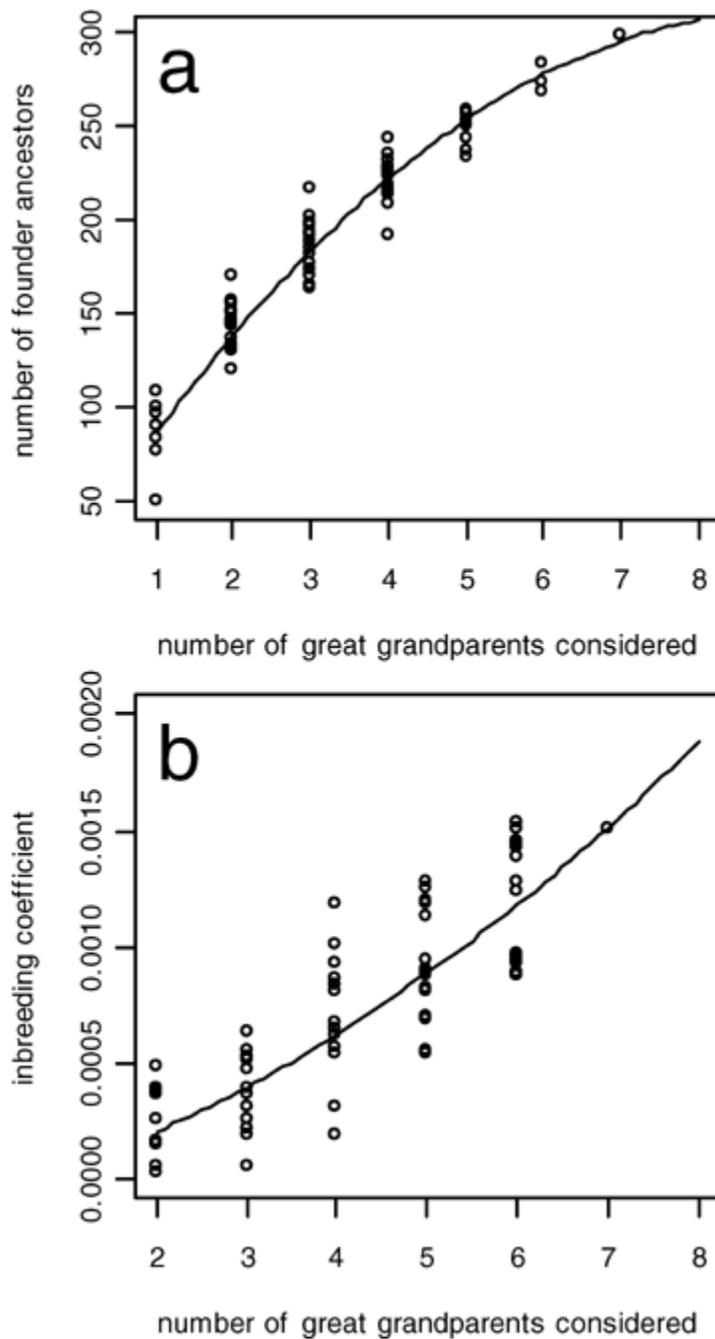
Wright, S. (1922) Coefficients of inbreeding and relationship. *Am Nat* **56**, 330–338.

•

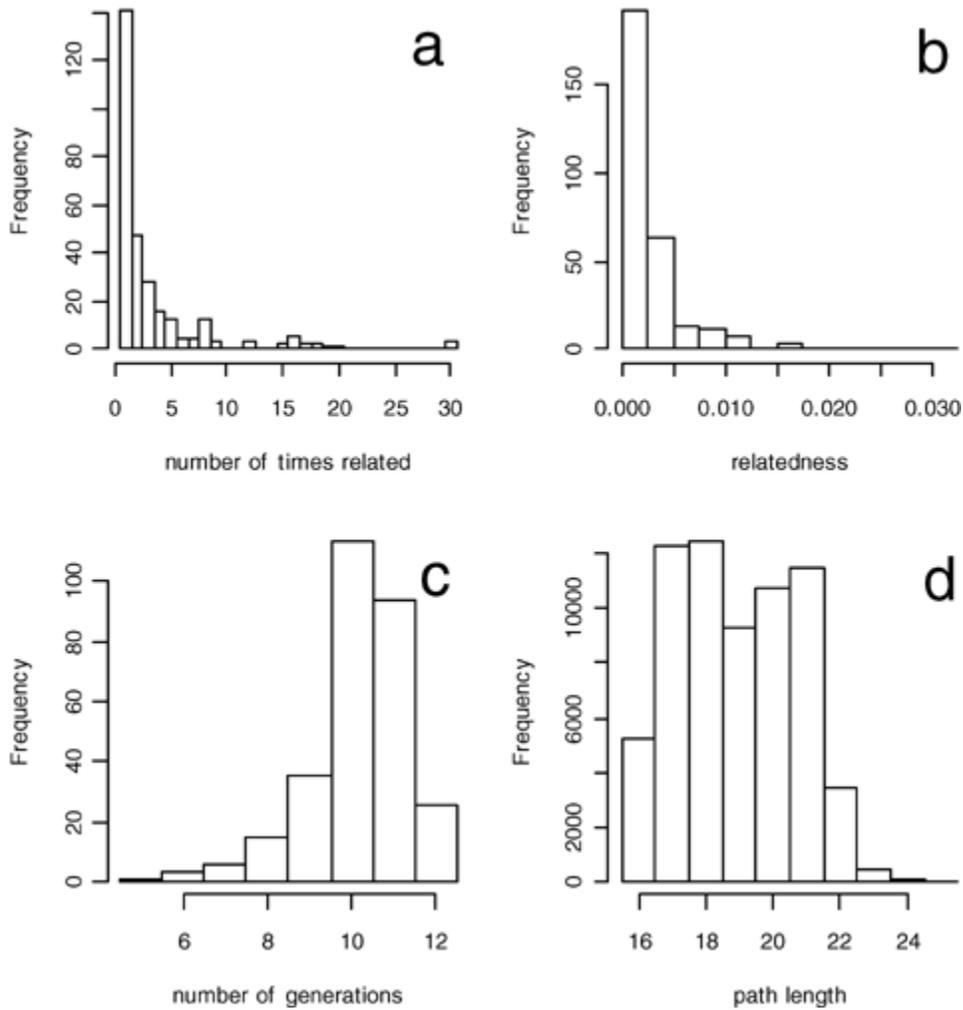
Wright, S. & McPhee, H. C. (1925) An approximate method of calculating coefficients of inbreeding and relationship from livestock pedigrees. *J Agr Res* **31**, 377–383.

## Figures and Tables

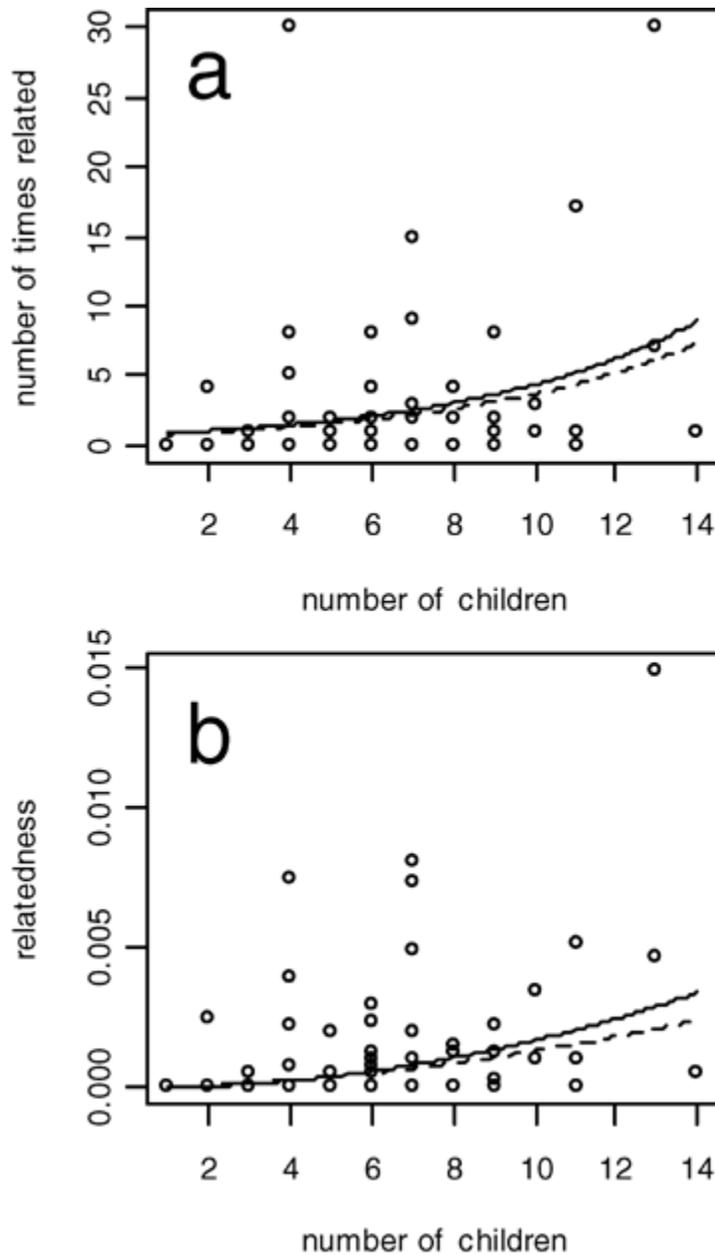
**Figure 1** a) Cumulative number of ancestors and b) cumulative inbreeding coefficient, as the random number of great grandparents considered is increased.



**Figure 2** Frequency distribution of **a)** my relatedness to each ancestor, **b)** the number of times I am related to each ancestor, **c)** number of generations between myself and founder ancestors and **d)** the path lengths (number of ancestors on a chain connecting my father and mother via a common ancestor).



**Figure 3** a) The relationship between the number of offspring an ancestor had and the number of times I am related to that ancestor. For this figure marriage date was taken as 1674.8. The dashed line is for when the outliers, where number of times related = 30, were left out. b) The relationship between the number of offspring an ancestor had and my relatedness to that ancestor. The dashed line is for when the outlier  $r = 0.01489$  is left out.



**Table 1** Number of emigrant couples married at the Cape, and a breakdown of the origin of the children these immigrants had, for five 30-year time windows.

Time	Parents		Children						
	Number of couples	Both Parents emigrants (%)	NL	DE	FR	non-European	Small European contributors	Unknown	UK
1657–1687	66	61 (92)	140	63	11	7.5	3.5	8.5	1
1688–1717	324	184 (57)	373	219.5	315	43.5	15.5	21	2
1718–1747	382	116 (30)	281	306.5	47.5	85	28	44.5	2
1748–1777	594	131 (22)	323.5	995.5	25	195	66	75	0
1778–1807	820	190 (23)	573.5	1222.5	52	291	111	88	26

Note: These data were extracted from Heese (1971). Note that this Table only considers couples that contain at least one emigrant, and that emigrants who were already married on arrival are not recorded here but their offspring are. The column headed "both parents emigrants" is a tally of couples in the previous column where both the husband and wife are emigrants.

**Table 2** Founder ancestors to whom I am related more than ten times.

Founder ancestor	n	g	relatedness	wedding date	country
VAN DER MERWE Willem Schalk	30	11.1	0.014892578	9.9.1668	NL
CLOETE Jacob	30	12.1	0.007446289	b. 1652	DE
RADERGRÖTZ Sophia	30	12.1	0.007446289		DE
PRÉVOST Charles	19	11	0.010009766	8.10.1673	FR
LE FÉBRE Marie	20	11	0.010498047		FR
BOTH Friedrich	18	10.8	0.010986328	21.6.1717	NL
KICKERS Maria	18	10.8	0.010986328		NL
VISSER Jan Coenraad	17	11.8	0.005126953	c 1653	NL
GERRITS Margaretha	17	11.8	0.005126953		NL
SNIJMAN Hans Christoffel	16	11.8	0.005004883	not married	NL
VAN PALICATTE Catharina	16	11.8	0.005004883		IN
DE SAVOYE Jacques	16	11.8	0.005004883	1665	FR
DU PONT Christine	16	11.8	0.005004883		FR
DES PRÉS Hercule	16	11.1	0.0078125	b. 1670	FR
D'ATIS Cecilia	16	11.1	0.0078125		FR
POTGIETER Harmen Jansen	15	10.9	0.008056641	8.5.1672	DE
FREDERIKS Isabella	15	10.9	0.008056641		NL
BURCHERDT Berndt	12	10.3	0.010253906	c 1690	DE
VERWEY Gysbert	12	11.4	0.004638672	b. 1668	NL
GANZEVANGER Catharina	12	11.4	0.004638672		NL

Note: n = number of times related and g is the average number of generations separating myself from this ancestor, country refers to the two letter code for the country of origin, under wedding dates b = before, c = circa. Names are arranged so that wives follow their husbands, except in two cases where the wife's parents had already been counted.

**Table 3** Founder ancestors to whom I am related by more than one percent.

<b>Founder ancestor</b>	<b>relatedness</b>	<b>n</b>	<b>g</b>
GAUKES Dirk Hendrik	0.03125	1	5
VAN MACAO Rosalyn	0.015625	1	6
KLEM Andreas	0.015625	1	6
FOSTER James	0.015625	1	6
VAN DER MERWE W. S.	0.014892578	30	11.1
LATEGAN Johann Hermann	0.01171875	4	8.5
BODENSTEIN Caspar	0.01171875	2	7.5
KICKERS Maria	0.010986328	18	10.8
BOTH Friedrich	0.010986328	18	10.8
LE FÉBRE Marie	0.010498047	20	11
BURCHERDT Berndt	0.010253906	12	10.3
PRÉVOST Charles	0.010009766	19	11
Note: n = number of times related and g is the average number of generations separating myself from this ancestor.			

**Table 4** Percentage contribution aggregated over different nationalities and in European and non-European contributions.

Classes of contributors	Nationality	This Study			Heese	
		Contribution	Scaled Contribution a	Scaled Contribution b	Scaled Contribution	Contribution
Three major contributors	Netherlands	30.8228	37.4518	37.4518	36.826	35.5
	German	22.5342	27.3806	27.3806	35.685	34.4
	French	21.7529	26.4313	26.4313	14.419	13.9
Lesser European contributors	Danish	0.1953	0.2373			
	Norwegian	0.0488	0.0593			
	Portuguese	0.3906	0.4746	0.7713	2.905	2.8
British	British	1.5625	1.8985	1.8985	2.697	2.6
Non-European	Chinese	1.5625	1.8985			
	Guinea	0.2197	0.2670			
	India	1.3794	1.6761			
	Madagaskar	0.0488	0.0593			
	van die Kaap	1.7822	2.1655	6.0664	7.469	7.2
Incomplete and unknown	Incomplete	17.1875				
	Unknown	0.5127				3.6

Note: The values under Heese refers to his calculations for Afrikaners as a whole (Heese, 1971). The scaled contributions for this study were obtained by dividing the contributions by 82.2998 (=100-17.1875-0.5127) and that of Heese's by 96.4 (=100-3.6). This corrects for the unknown and incomplete lineages and makes it possible to compare my results to that of Heese (1971). Such a comparison makes the implicit assumption that the unaccounted lineages will essentially be the same as the recorded lineages. Scaled contribution 'a' gives the scaled values for all the nationalities I recorded, whereas scaled contribution 'b' groups together my data into the classes that Heese (1971) recorded.

**Table 5** My non-European founder ancestors.

<b>Origin</b>	<b>Name</b>	<b>relatedness</b>	<b>n</b>	<b>g</b>
Chinese	VAN MACAO Rosalyn	0.01563	1	6
Guinea	VAN DIE KAAP Lijsbeth SANDERS	0.00171	4	11.3
	VAN GUINEE Anna	0.00024	1	12
	VAN GUINEE Evert	0.00024	1	12
India	VAN BENGAL Catharina OPKLIM	0.00098	1	10
	VAN BENGAL Magteld Maria Cornelisse	0.00098	1	10
	VAN MALABAR Helena	0.00098	1	10
	VAN MALABAR\COROMANDEL Cath.	0.00146	4	11.5
	VAN NEGAPATNAM Maria	0.00439	8	10.9
	VAN PALICATTE Catharina	0.00500	16	11.8
Madagascar	VAN MADAGASKAR Diana	0.00049	1	11
van die Kaap	Unknown woman, mother of Stoltz	0.00146	2	10.5
	VAN DER HEYDE VAN DIE KAAP Anna	0.00146	2	10.5
	VAN DIE KAAP Ansela	0.00122	3	11.3
	VAN DIE KAAP Catharina	0.00049	1	11
	VAN DIE KAAP Juliana Constant	0.00146	2	10.5
	VAN DIE KAAP Margaretha	0.00049	1	11
	VAN DIE KAAP Maria	0.00098	1	10
	VAN DIE KAAP Maria BASTIAANSZ	0.00195	1	9
	VAN DIE KAAP Maria LOZEE	0.00244	4	10.8
	VAN DIE KAAP Sus. Marg. FYNTON	0.00391	1	8
	VRYMAN Catharina	0.00195	3	10.7

Note: n = number of times related and g average number of generations separating myself from these ancestors.

**Table 6** Summary of the statistical models fitted to fitness.

Explanatory variable	Predictor variables	Including high outliers		Excluding high outliers	
		coefficient	P	coefficient	P
number of times related	intercept	111.01612	0.0016	-0.39864	0.427
	number of children	0.17879	0.0008	0.17151	0.006
	marriage date	-0.06647	0.0010		
relatedness	intercept	-0.00099	0.8932	0.00337	0.6373
	number of children	0.0042271	0.0003	0.00329	0.0038

Note: The minimum adequate generalized linear model and linear model that were respectively fitted to explain the number of times related and the relatedness as a function of number of children. For the former the antilog of the predictor needs to be taken and for the latter the square. For the GLM the data was over-dispersed, and the quasi-Poisson option was used in R (RTeam 2005). The adjusted  $R^2$  for the latter was 0.1804.