Am J Hum Genet 32:445–452, 1980

Kinship Structure and Heterozygosity on Tristan da Cunha

E. A. THOMPSON[1] AND D. F. ROBERTS[2]

SUMMARY

Although there are many factors influencing the genotype proportions in a small population [1], a heterozygote excess, caused by avoidance of incestuous matings, will often be the expected observation. In the population of the small South-Atlantic island of Tristan da Cunha, such an excess is observed [2], and we have investigated its origin, partitioning the population by sex and by generation. Although for such a small population the test for random pairing of genes [3, 4] is not sufficiently powerful to produce significant results, the pattern of homozygote deficiency is suggestive of avoidance of close matings. We have also investigated the effect of current nuclear family structure and family-size distribution.

INTRODUCTION

The effect of a population's mating structure upon its genotypic structure has been of long-standing interest. Wright [5] first considered matings between related individuals in an infinite population. Using his "fixation index," he analyzed the excess homozygosity that would arise over the Hardy-Weinberg prediction. When mating occurs at random in a population, offspring genotypes are expected to arise in Hardy-Weinberg proportions. In a large population, mates will, by virtue of social and geographical proximity, tend to be more closely related to each other than two randomly chosen individuals. An overall excess of homozygotes is, therefore, expected, and where the mating structure results from geographical or social subdivision of the population, this excess may be measured in terms of the Wahlund variance (see [6], p 54, for example). There are, however, many other factors that influence the genotypic structure of a population, and even in a large population, an excess of

Received July 11, 1979; revised September 25, 1979.

[1] King's College, Cambridge, England.

[2] Department of Human Genetics, University of Newcastle, England.

heterozygotes may be observed. Smith [1] has considered theoretically several possible explanations for such an observation.

In a small population, however, no such explanation is required. Often, because of taboos on between-sibs or uncle-niece matings, for example, mates are less closely related than two randomly chosen members of the population. In this case, an excess of heterozygotes is the expected result, not an anomalous occurrence. Roberts [2] computed the expected excess of *homozygotes* on the basis of kinship between mates. Neel [7] also expected to see such an excess among American Indians, and Jacquard and Bengtsson [8] have considered theoretically the expected level of such homozygote excess. This excess is, however, relative to the founding population, and to the (unknown) allele frequencies which are obtained within the founder group ([6], p 56). Within the current population, the relevant factor is not the absolute value of kinship between mates, relative to a founding gene pool, but this kinship's value relative to that between random members of the population.

POPULATION THEORY

Jacquard ([9], p 221) introduces a coefficient δ which measures deviation from random mating in terms of its expected effect on genotypic proportions. The coefficient is defined by the equation

$$G_M = (1 - \delta)\, G_P + \delta G_H \,, \tag{1}$$

where G_M is the genotypic structure of a gene pair from mates (i.e., the expected genotypic structure of offspring produced); G_P, the panmictic Hardy-Weinberg structure; and G_H, the genotypic structure of a completely homozygous population with the same allele frequencies. Where mating patterns are defined by genealogical relationship, and not by genotype or phenotype, a linear relationship of the form of equation (1) will always hold. If kinship between two random members of the population, relative to some defined ancestral genealogy, is α, while that between mates is α^*, then equation (1) implies $\alpha^* = (1 - \delta)\, \alpha + \delta.1$ or

$$(1 - \alpha^*) = (1 - \alpha)\,(1 - \delta) \,. \tag{2}$$

This equation has an obvious parallel in the classical equation of [10]:

$$(1 - F_{IT}) = (1 - F_{ST})\,(1 - F_{IS}) \,, \tag{3}$$

and can be so interpreted if we consider I as an individual resulting from a mating; S, the particular subpopulation under consideration; and T, the totality of all populations with the same underlying ancestral structure. The coefficient δ may thus be interpreted as a correlation between uniting gametes relative to the current genotypic structure of the population and may be positive or negative: α and α^* are necessarily positive.

Jacquard [9] has computed δ for a discrete-generation population of size N, for several cases of prohibition of matings between relatives. These provide a basis for comparison with avoidance patterns in natural populations. In any population where there is a net avoidance of matings between close relatives, and, hence, in any small natural population such as that of Tristan da Cunha, δ will be negative.

TEST OF HARDY-WEINBERG PROPORTIONS

For any small population of interest, we may observe the current phenotypic or genotypic frequencies at particular loci. We shall assume that the whole population is surveyed and shall also restrict attention to loci with two codominant alleles, since for such loci any excess or deficiency of homozygotes is most readily assessed. Since our population is not a sample from an infinite population, nor in any sense a random sample as it is one interrelated whole, what is the purpose of estimation and testing? We do not wish to estimate an allele frequency; the allele frequency *is* that observed. We may, however, wish to consider whether the observed number of homozygotes is in accord with that expected under random mating. That is, given the $2n$ homologous genes of the population, paired at random into n diploid individuals, what is the probability of obtaining a number of homozygotes as small (or as large) as that observed? We are, thus, in a situation in which the conditional test procedure of Haldane [3] (see also [11]) is applicable (even though that procedure was based on a sample from an infinite population). Levene [4] has provided the factorial moments of the numbers of each genotype under the above random pairing, while Shanbhag and Rao [12] have considered the asymptotic properties of Haldane's test and extended it to multiple alleles. In the case of two codominant alleles, it is, of course, only necessary to consider one genotype, since, given the numbers of each allele, the numbers of other genotypes are then determined.

TRISTAN DA CUNHA KINSHIP STRUCTURE

For the small and isolated population of Tristan da Cunha, we have the complete genealogy back to the few original founders as well as genetic data for almost the complete 1961 population [13]. The population, therefore, provides an ideal situation for the study of mating preference in a small population; some particular cases of limitation in the choice of marriage partner, and the resulting inbreeding levels, have been discussed previously [2]. Also in [2], levels of mean inbreeding, and random kinship over the whole population at different points in history, were considered. However, it proved difficult to see any parallel between them, largely because of the population's heterogeneity caused by grouping together disparate generations.

Since we wish to compare inbreeding of offspring with kinship between potential parents, we partitioned the population at each date into two generations. The offspring generation at each date is considered to be individuals from birth to age 20. The parent generation consists of males aged 21 to 60 and females 21 to 40. From the complete genealogy of the relevant individuals we computed: (i) the average inbreeding coefficient in each offspring generation; (ii) the average kinship coefficient over all pairs in the parental generation; (iii) the average kinship coefficient between all nonsib pairs; and (iv) the average kinship coefficient excluding all pairs with a value of this coefficient greater than 1/8. Since potential parent couples are of primary interest, we have also computed the kinship coefficients (ii), (iii), and (iv) for male-female pairs only. The results are given in table 1.

Inbreeding levels among offspring have been consistently smaller than mean kinship in the parent generation, indicating a net avoidance of inbreeding throughout the

TABLE 1

HISTORY OF THE TRISTAN DA CUNHA POPULATION

Year	No. individuals: Adults: Male	Adults: Female	Children	Mean parental kinships × 1,000: Overall: (ii)*	Overall: (iii)	Overall: (iv)	Male-female pairs: (ii)	Male-female pairs: (iii)	Male-female pairs: (iv)	Offspring inbreeding × 1,000 (i)*	δ-values from kinship structure: All pairs × 1,000 (4)†	δ-values from kinship structure: Male-female pairs × 1,000 (4)†	Theoretical δ-values‡ × 1,000: (5)†	Theoretical δ-values‡ × 1,000: (6)†
1840	6	7	...	14	5	0	12	0	0	...	...	...	...	...
1850	8	7	...	51	1	0	54	0	0	...	...	...	...	...
1860	8	7	20	47	18	14	45	16	11	23	−25	−23	−28	−100
1870	13	10§	37	69	27	26	69	26	26	23	−49	−49	−14	−37
1880	18§	5	63	56	34	25	57	27	16	40	−17	−18	−8	−19
1890	7	6	35	60	40	36	62	41	36	33	−29	−31	−15	−40
1900	14	11	37	64	44	39	73	43	38	34	−32	−42	−14	−37
1910	20	13	53	68	47	32	65	49	35	31	−40	−36	−10	−23
1920	23	15	64	67	58	44	63	59	43	46	−20	−15	−10	−18
1930	40	21	80	72	61	51	73	61	51	70	−2	−3	−6	−14
1940	44	24	100	73	61	53	76	63	55	74	+1	−2	−5	−11
1950	63	33	107	71	64	55	69	64	54	61	−11	−9	−5	−10
1960	72	40	100	70	66	58	70	66	58	54	−17	−17	−5	−11

* See TRISTAN DA CUNHA KINSHIP STRUCTURE in text for definitions of inbreeding and kinships (i) to (iv).

† Nos. refer to equations in text.

‡ See text.

§ One individual, unrelated to any other inhabitant and having no descendants, is omitted here.

island's history. It seems, however, that this net avoidance is caused only by avoidance of incestuous matings. Although the inbreeding level is usually below the kinship level between nonsib pairs, it is close to it and to kinship between pairs for whom this is less than 1/8 (excluding uncle-niece, for example). This has been so since 1870, the only exception being a significantly higher inbreeding rate in the 1930s. The period 1920–1940 was seemingly characterized by a rather different mating pattern. (Before 1870, the numbers of individuals involved are too small for any conclusions to be drawn.) The overall kinships scarcely differ from those between male-female pairs, showing that the random factor of an individual's sex has not materially affected mate availability.

We computed the expected levels of heterozygote excess under the given kinship structure. From equation (2)

$$\delta = (\alpha^* - \alpha)/(1 - \alpha) \ , \tag{4}$$

where α^* is the inbreeding coefficient of offspring, and α is the mean kinship between random pairs of the parental generation. Values of δ are, therefore, also given in table 1, both for random pairs and random male-female pairs in the parent generation, and may be compared with theoretical values given by [9]. For a discrete generation population of constant size n, in which sib mating is excluded, but mating is otherwise at random,

$$\delta = -1/2(n - 2) \ ; \tag{5}$$

while if first-cousin mating is also excluded,

$$\delta = -1/(n - 10) \ . \tag{6}$$

These values are, therefore, also given in table 1, with the current parental value of n taken as $n = 4(1/n_m + 1/n_f)^{-1}$ to adjust for the different numbers of males (n_m) and females (n_f). These values, of course, provide only a general indication; the derivation of equations (5) and (6) involved assumptions of Poisson family-size distribution, etc., which are, at best, an approximation in a natural population.

We again see that avoidance of inbreeding in this population is approximately to the extent of sib avoidance; there are, of course, many first-cousin marriages in the history of Tristan da Cunha, but apparently neither significantly more nor less than expected by chance. The high inbreeding levels of individuals who were children in 1930 and in 1940 again stand out, but in 1950 and in 1960, the avoidance has again increased. In fact, the increase in avoidance in these recent years appears to be greater than implied by sib-avoidance alone for a population of this size. However, in this larger population, the large variance of sibship size becomes an important factor, and it accounts for the discrepancy between the next to the last column of table 1 and the two preceding columns.

KINSHIP PATTERNS AND GENOTYPE FREQUENCIES

There are three loci, *MN*, *Rh*-C, and *Rh*-E, each having two codominant alleles, for which almost the complete 1961 population of Tristan da Cunha has been sampled. The number of individuals of homozygote types *MN*, *CC*, and *EE* and the numbers of *M*,

C, and *E* alleles in the population are given in table 2. These loci were chosen according to [2], although the two *Rhesus* loci are not independent—any *CC* individual in this population is also *ee*, having two R_1 or R' alleles. The population sampled has again been partitioned by sex and generation: into male and female, and into those from birth to age 20 in 1960, aged 21–40 (females), 21–60 (males), and those above the parental age limit. The offspring and parent generations thus correspond to the final line of table 1. The total numbers of each genotype differ slightly from those of [2]; a few individuals were typed subsequent to the initial collection of data, and these are included here.

Also given in table 2 are the mean and standard deviation (SD) of the number of individuals of the specified homozygote type, conditional on the observed numbers of each allele, under the hypothesis of random pairing of the genes. As noted by [2], there is, overall, for each of the three loci, a deficiency of homozygotes. However, only for the *MN* locus does the deficiency exceed 1SD and even in this case, it is less than 2. Thus, a hypothesis of random pairing could not be rejected on the basis of these data. It is, nonetheless, of interest to investigate further the source of this deficiency. A true deficiency may be obscured by pooling together different populations; thus following our previous discussion, we have considered the excess or deficiency of homozygotes in each subclass of the 1961 population, partitioned by age and sex. Again, there is in no class a significant deficiency of homozygotes, yet in view of the small numbers of individuals involved, it is impressive that so large a proportion of the classes should show some deficiency. This provides some evidence for the avoidance of inbreeding in this population.

In addition, table 2 gives the expected numbers of each allele and of the homozygote type for: (a) random offspring of the parental generation; (b) random offspring of the actual parents of the current offspring generation mated at random; and (c) random offspring of each actual parent pair, one such random offspring being generated for each actual offspring that exists in the current offspring generation. Under random mating, the only deviations from Hardy-Weinberg proportions expected among the offspring arise from differences in male and female allele frequencies, causing a heterozygote excess. Yet even where the male-female parental difference is substantial (at the *Rh*-E locus, for example), the expected homozygote deficiency due to this factor is only .14, compared with a SD of 2.34 under the random-pairing hypothesis. The differences between the parental sets defined by age (a), and parental sets defined by actual parenthood (b), are very small, and in the case of the *Rhesus* loci, these are both very close to the expectations for the actual current parent pairs (with multiple parents counted the appropriate number of times). Thus, our definition of the parental set which we have used to study the historical data does closely approximate the true parental generation in 1960 at least, and for the *Rhesus* loci, the distorting effect of nuclear family structure is also negligible.

The data for the *MN* locus are of greater interest. In the offspring generation, we observed a very slight homozygote (*MM*) deficiency, this being composed of a slight excess among male offspring overcompensated by a deficiency among female offspring. Among the parents, there is a homozygote deficiency in all classes, the numbers being in each case between 1 and 2 SDs below those predicted by random

TABLE 2

ANALYSIS OF THE 1961 TRISTAN DA CUNHA POPULATION GENOTYPE DATA

	No. individuals	*MN* locus				*Rh*-C locus				*Rh*-E locus			
		$n(MM)$	$n(M)$	$\mathbb{E}^*(MM)$	$\sigma^*(MM)$	$n(CC)$	$n(C)$	$\mathbb{E}(CC)$	$\sigma(CC)$	$n(EE)$	$n(E)$	$\mathbb{E}(EE)$	$\sigma(EE)$
Males born:													
Since 1960	4	2	5	†	†	0	1	†	†	2	4	†	†
1940–1960	42	7	34	6.76‡	1.57	2	22	2.78	1.26	8	36	7.59‡	1.60
1900–1939	72	15	70	16.89	1.26	5	44	6.62	2.57	8	51	8.92	2.99
Before 1900	9	3	10	†	†	0	6	†	†	1	6	†	†
Females born:													
Since 1960	6	1	5	†	†	0	3	†	†	0	3	†	†
1940–1960	52	10	48	10.95	1.81	7	33	5.13‡	1.55	7	42	8.36	2.88
1920–1939	42	7	40	9.40	1.61	3	21	2.53‡	1.21	7	38	8.47	1.62
Before 1920	34	4	29	6.06	1.44	1	19	2.55	1.18	4	27	5.24	1.40
Total:													
Males	127	27	119	§	§	7	73	§	§	19	97	§	§
Females	134	22	122	§	§	11	76	§	§	18	110	§	§
Offspring	94	17	82	17.76	2.39	9	55	7.94	2.01	15	78	16.06	2.35
Parents	114	22	110	26.41	2.67	8	65	9.16	2.19	15	89	17.25	2.55
Grand Total	261	49	241	55.51‖	4.00	18	149	21.16	3.32	37	207	40.92	3.90
Expectation for 94 random offspring of:													
(a) Parental set		21.2	90	21.4	2.57	7.0	52	7.09	1.94	15.1	76	15.24	2.34
(b) Actual parents randomly mated		20.2	89	20.9	2.75	6.8	51	6.82	1.91	14.8	75	14.84	2.33
(c) Actual parent couples		25.1	92	22.4‡	2.29	7.3	52	7.09‡	1.94	13.8	77	15.65	2.33

* $\mathbb{E}$ = expectation; σ = standard deviation.

† Nos. too small for meaningful computation.

‡ Observed no. homozygotes *larger* than expected under random pairing hypothesis.

§ Not computed, since a nonhomogeneous group with respect to age.

‖ Computed, although a nonhomogeneous group with respect to age, for comparison with [2].

pairing. Random mating among the parental generation predicts (necessarily) a homozygote deficiency among the offspring; but in this case, the predicted deviation from random pairing is very small, the male and female parental allele frequencies being almost equal. On generating random offspring to the actual parent couples, we predict, however, a homozygote *excess* [see (c) of table 2]. This excess is again within the bounds imposed by a random-pairing hypothesis; yet it is clear that, for this locus at least, the nuclear family structure of the current generation has substantial effect. It would be of interest to examine the sibship structure of the parent generation, where the deficiency of homozygotes is far more apparent, to see to what extent this is attributable to nonrandom mating among *their* parents. However, the genetic data for these parents are unavailable. They may be estimated from their offspring types, yet this procedure loses power, rendering the final results again insignificant, although again suggestive.

To relate this homozygote deficiency to kinship structure, we have also analyzed the historical data. Over most of this population's history, the kinship and inbreeding structure closely approximates that of a population mating at random, apart from avoidance of sib, half-sib, parent-offspring, and uncle-niece matings. Although a possible major factor in some small populations, the sex of individuals seems to have little influence in affecting the kinship structure (by mate availability), either historically or within the current population.

ACKNOWLEDGMENT

We are grateful to Mrs. M. Hindmarsh for doing the computing which forms the basis of table 1.

REFERENCES

1. SMITH CAB: Measures of homozygosity and inbreeding in populations. *Ann Hum Genet* 37:377–391, 1974
2. ROBERTS DF: The development of inbreeding in an island population. *Ciencia e Cultura* 19:78–84, 1967; reprinted *Yearbook of Physical Anthropology* 15:150–160, 1967
3. HALDANE JBS: An exact test for randomness of mating. *J Genet* 52:631–635, 1954
4. LEVENE H: On a matching problem arising in genetics. *Ann Math Stat* 20:91–94, 1949
5. WRIGHT S: Systems of mating. *Genetics* 6:111–178, 1921
6. CROW JF, KIMURA M: *An Introduction to Population Genetics Theory*. New York, Harper & Row, 1970
7. NEEL JV, LAYRISSE M, SALZANO FM: Man in the tropics: the Yanomama Indians, in *Population Structure and Human Variation*, edited by HARRISON GA, International Biological Programme, vol. II, Cambridge, Cambridge Univ Press, 1977
8. JACQUARD A, BENGTSSON BO: Loi de distribution des homozygotes dans une population. *Population* 31:64–71, 1976
9. JACQUARD A: *The Genetic Structure of Populations*. New York, Springer-Verlag, 1974
10. WRIGHT S: The genetical structure of populations. *Ann Eugen (Lond)* 15:323–354, 1951
11. EMIGH TH, KEMPTHORNE O: A note on goodness of fit to Hardy-Weinberg structure. *Am J Hum Genet* 27:778–783, 1975
12. SHANBHAG DN, RAO DC: Haldane's exact test for randomness of mating. *J Genet* 61:169–176, 1974
13. ROBERTS DF: The demography of Tristan da Cunha. *Population Studies* 25:465–479, 1971