How large was the founding population of Darwin's finches?

VLADIMIR VINCEK[1], COLM O'HUIGIN[2], YOKO SATTA[3], NAOYUKI TAKAHATA[3], PETER T. BOAG[4], PETER R. GRANT[5], B. ROSEMARY GRANT[4] AND JAN KLEIN[1,2]

[1] *Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, FL 33136, USA*
[2] *Max-Plank-Institut für Biologie, Abteilung Immungenetik, Corrensstr. 42, D-72076 Tübingen, Germany*
[3] *Coordination Center for Research and Education, The Graduate University for Advanced Studies, Hayama, Kanagawa 240-01, Japan*
[4] *Department of Biology, Queen's University, Kingston, Ontario, Canada K7L 3N6*
[5] *Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544-1003, USA*

SUMMARY

A key assumption of many allopatric speciation models is that evolution in peripheral or isolated populations is facilitated by drastic reductions in population size. Population bottlenecks are believed to lead to rapid changes in gene frequencies through genetic drift, to facilitate rapid emergence of novel phenotypes, and to enhance reproductive isolation via genetic revolutions. For such effects to occur, founding populations must be very small, and remain small for some time after founding. This assumption has, however, rarely been tested in nature. One approach is to exploit the polymorphism of the major histocompatibility complex (*Mhc*) to obtain information about the founding population. Here, we use the *Mhc* polymorphism to estimate the size of the founding population of Darwin's finches in the Galápagos Archipelago. The results indicate that the population could not have been smaller than 30 individuals.

1. INTRODUCTION

Speciation involves genetic changes that differentiate gene pools derived from an ancestral pool and ultimately produce a reproductive barrier between the emerging species (Dobzhansky 1937; Mayr 1942, 1963; Otte & Endler 1989). The changes have traditionally been studied by three principal methods: by following genes in artificial populations over a certain number of generations (reviewed in Rice & Hostert 1993); by cross-breeding closely related species to reveal the underlying genetic differences and inferring the changes responsible for them (e.g. Coyne *et al*. 1994; Wu *et al*. 1995); or by developing theoretical models and examining their feasibility analytically or by computer simulation (e.g. Templeton 1981). Recently, however, a fourth method has become available by which a population's demographic history is inferred from characteristics of extant gene pools (Klein *et al*. 1993*a*). This new approach has been made possible by two major innovations: the formulation of the gene genealogy (coalescence) theory (Kingman 1982*a*, *b*; Tavaré 1984; Watterson 1984; Hudson 1990) and the development of rapid methods of gene amplification and sequencing (Saiki *et al*. 1988). Genes that display long-term balanced polymorphism often have coalescence times longer than the life-span of a species (Takahata 1990; Takahata & Nei 1990). Such genes can then be used to make inferences about past speciation events, especially about the size of the founding population. A particularly rich source of genealogical information has been found in the genes of the major histocompatibility complex, *Mhc* (Klein *et al*. 1990*a*; Klein *et al*. 1993*b*).

The *Mhc* is a group of loci specifying cell-surface receptors for peptides derived from self and nonself proteins (Klein 1986; Srivastava *et al*. 1991). The *Mhc* genes are under balancing selection and are transmitted from ancestral to descendant species as conserved allelic lineages (Klein 1987; Klein *et al*. 1993*a*). This property has been exploited to estimate the size of the founding populations of *Homo sapiens* (Klein *et al*. 1990*a*, *b*, 1993*a*) and of cichlid fishes in Great East African Lakes (D. Klein *et al*. 1993; Ono *et al*. 1993).

One group of speciation hypotheses is based on the assumption that the founding populations are generally very small (Templeton 1980; Carson & Templeton 1984; Giddings *et al*. 1989; Hollocher 1996). Populations are believed to pass through one or more bottleneck phases which create conditions for rapid remodelling of the genotype and fixation of genes responsible for reproductive isolation of the emerging species. According to the hypothesis new species arise from a single inseminated female, or at most from very few individuals. This situation is believed to occur during colonization of geographically distant places and is exemplified by the radiation of Darwin's finches on the Galápagos Islands.

The Galápagos Archipelago is a group of 13 large and six smaller islands some 960 km west of the coast of

Ecuador (Jackson 1993; Simkin 1984). Although recently obtained evidence indicates the presence in this region of a series of sunken islands which may have formed more than 10 million years ago (MYA) (Christie *et al*. 1992), most existing islands were formed between 0.5 and 5 MYA (Cox 1983; White *et al*. 1993; Geist 1996).

At some point in the past, seed-eating finches reached the Galápagos Islands, presumably flying from the western coast of South America (Lack 1947; Grant 1986; Grant & Grant 1989). On arrival they underwent adaptive radiation filling the different ecological niches available on the islands and produced 13 extant species of Darwin's finches. The species differ in feeding habits, beak size and shape and body size. To estimate the size of the finch population that originally colonized the islands we identified the *Mhc* class II genes of Darwin's finches and determined the extent of their variation.

2. MATERIALS AND METHODS

DNA was extracted from small blood samples collected from living birds captured on the Galápagos Islands. The primers used for polymerase chain reaction (PCR) amplification (HOPE1 5′GAAAGCTCGAGTGTCACTTCACGAACGGC3′ and HOPE2 5′GGGTGACAATCCGGTAGTTGTGCCGGCAG3′) were synthesized by Keystone Laboratories, Inc., Menlo Park, CA. The PCR was carried out as described earlier (Vincek *et al*. 1995). The PCR products were purified by electrophoresis in low-melting point agarose (Gibco BRL) and subcloned in the pUC18 plasmid vector using Pharmacia Sure Clone kit. Double-stranded DNA was isolated from individual subclones using Qiagen plasmid kits and sequenced by the dideoxy chain-termination method (Sanger *et al*. 1977) using version 2.0 sequenase kit (US Biochemicals) and [α^{35}S]dATP.

Substitutional distances were measured using Kimura's two-parameter method (Kimura 1980), and dendrograms were constructed by the neighbour-joining algorithm (Saitou & Nei 1987) and the unweighted pair-group method with arithmetic mean (UPGMA; Sokal & Michener 1958). To simulate the effect of the founder population size on the maintenance of alleles a computer program was developed, based on simple assumptions. A parental population (gene pool) was assumed to have existed on the continent, from which a random sample was drawn to form the founding population. For want of a more appropriate model, the parental population was assumed to have contained the same number of alleles (35) and the same allelic frequencies as the *HLA-DRB*1 locus in the Caucasoid human population (see Imanishi *et al*. 1991). From this population, $2N_b$ genes were drawn randomly and paired. Each pair was then evaluated for homo- and heterozygosity and the two randomly chosen genes were passed onto the next generation with the probability of one (heterozygotes) or 1-s (homozygotes), where s was the selection coefficient. The random drawing of genes for the t^{th} generation after the founding event continued until either $2N_b$ or $2N_b\,(1+r)^t$ (depending on the model) size of gene pool was reached (where r was the growth rate). This process was repeated for ten generations (model B) or until the population size of 1000 individuals was reached (models A, C–E), at which time the number of alleles retained in the population was determined. The entire simulation was repeated 1000 times, and the mean number of retained alleles as well as the standard error of the mean was calculated and plotted on the ordinate of a graph against the chosen N_b value on the abscissa. A different N_b value was then chosen and the entire procedure repeated. In this way, curves were obtained indicating the number of retained alleles for different $2N_b$ sizes and different simulation parameters. Mutations were not introduced into the simulation model because we were concerned with allelic lineages that were present in the ancestral populations at different times in the past and not with alleles produced by mutations after the founding event.

3. RESULTS AND DISCUSSION

To identify the *Mhc* genes of Darwin's finches, we designed PCR primers on the basis of *Mhc* class II sequences of the Bengalese finch (*Lonchura striata*; Vincek *et al*. 1995) and domestic fowl (Wittzell *et al*. 1994). The oligonucleotides were complementary to the 5′ and 3′ ends of exon 2, to regions that are relatively conserved among various *Mhc* class II gene sequences. The products obtained by PCR amplification using these primers were 173 bp long and encompassed the most polymorphic part of *Mhc* class II genes (Klein & Figueroa 1986). We were able to amplify the DNA from 20 birds representing eight species: *Cactospiza pallida* (*Capa*, 1), *Certhidea olivacea* (*Ceol*, 1), *Geospiza cornirostris* (*Geco*, 2), *Geospiza fortis* (*Gefo*, 6), *Geospiza fuliginosa* (*Gefu*, 2), *Geospiza magnirostris* (*Gema*, 3), *Geospiza scandens*, (*Gesc*, 3), and *Platyspiza crassirostris* (*Plcr*, 2). (Here the letters in parentheses are the species *Mhc* symbols according to the nomenclature system proposed by Klein *et al*. 1990*c*; the numbers indicate the number of individuals tested.) The amplification products were cloned and sequenced. Altogether we obtained 103 sequences of which 60 were different from one another (the sequences have been deposited in the EMBL databank and can be retrieved under the accession codes Z74411–Z74470). Identical sequences were found mostly in individuals of the same species (33/103) and fewer in individuals of different species (10/103). Based on their similarity, the sequences could be arranged into four groups. In addition to the greater overall similarity, sequences within each group shared sequence motifs – short stretches of residues present in most sequences of a given group. The motifs stand out particularly in the amino acid sequences translated from the nucleotide sequences. The four groups are clearly separated in the dendogram based on the nucleotide sequences (figure 1). Each group, except group four, contains sequences from most of the species tested. Groups 1–4 contain 21, 18, 13 and 8 sequences, respectively.

On the basis of the following two observations, we conclude that the four groups represent different loci. First, whenever a similar situation has been encountered in other vertebrate species, the groups could ultimately be shown to constitute different loci. A good example is provided by the cichlid species of the Great East African lakes (D. Klein *et al*. 1993; Ono *et al*. 1993), in which the *Mhc* sequences were initially classified into 13 groups on the basis of similarities in the order of those observed in Darwin's finches, and the groups were later shown to represent distinct loci (E. Malaga *et al*. unpublished data). Second, several

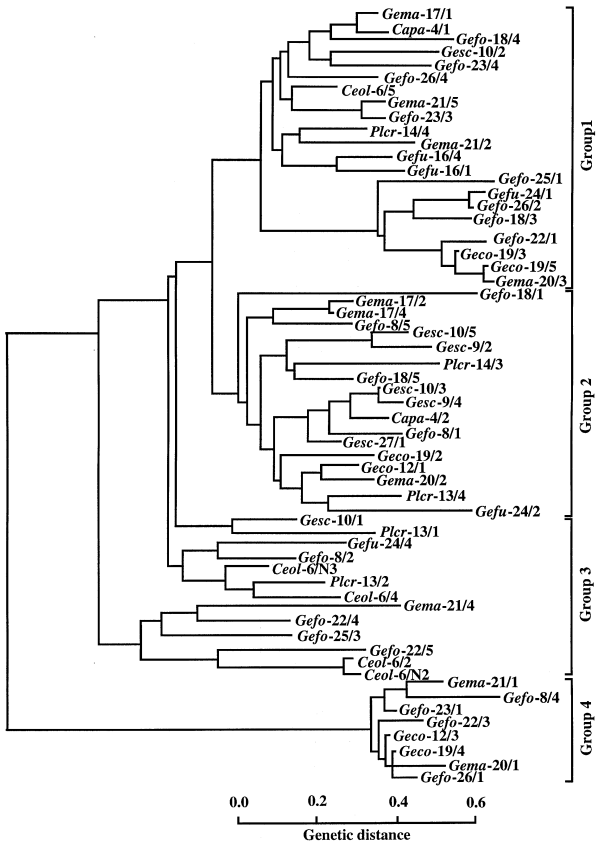

Figure 1. Phylogenetic tree of *Mhc* class II *B* exon 2 sequences of Darwin's finches. The tree is based on substitutional distances measured by the two-parameter method at all 173 nucleotide sites. It was constructed by the neighbour-joining method. The four clades, presumably representing distinct loci, are designated as Groups 1–4. Species abbreviations: Capa, *Cactospiza pallida*; Ceol, *Certhidea olivacea*; Geco, *Geospiza cornirostris*; Gefo, *Geospiza fortis*; Gefu, *Geospiza fuliginosa*; Gema, *Geospiza magnirostris*; Gesc, *Geospiza scandens*; Plcr, *Platyspiza crassirostris*. The numbers indicate individual birds/individual sequences.

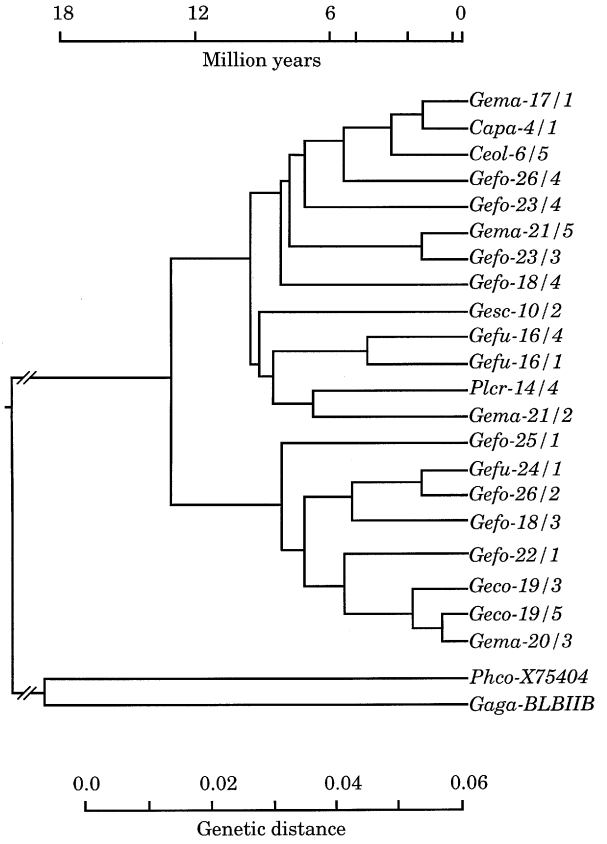

Figure 2. UPGMA tree of *Mhc* Class II *B* Group 1 sequences of Darwin's finches, domestic fowl and pheasant (see figure 1). The ticks on the time scale indicate the estimated time at which the ancestors of Darwin's finches reached the Galápagos Islands. *Gaga*, *Gallus gallus* (domestic fowl; sequence from Wittzell *et al*. 1994); *Phco*, *Phasianus colchicus* (pheasant; sequence from Zoorob *et al*. 1990); other symbols are explained in figure 1.

individual birds were shown to possess more than two different sequences. In all such cases, the multiple sequences of a given bird fell into more than one group.

Similarly, two observations support the conclusion that at least Group 1 represents a set of alleles at a single locus. First, no more than two sequences of this group were found in any individual bird. Second, sequences of this group share characteristic motifs; sharing of motifs between sequences of different loci has not been observed in any vertebrate species. These two arguments apply also to Groups 2 and 4; Group three may represent two loci. Group 1 consists of 21 sequences which we conclude are all derived from alleles at a single locus. To estimate the size of the founding population of Darwin's finches, it is necessary to know how many of these alleles (allelic lineages) were present in the founding population. To obtain such an estimate, we constructed a dendogram of Group one using the UPGMA (figure 2), which is more suitable for estimating coalescence time than the neighbour-joining method (Nei 1987). Next, we obtained a time scale for the UPGMA tree by comparing the *Mhc* class II sequences of the pheasant (Zoorob *et al*. 1990) and the domestic fowl (Wittzell *et al*. 1994), two species which are believed to have diverged 20 MYA (Helm-Bychowski & Wilson 1986). The comparison yielded an evolutionary rate of 3.9×10^{-9} nonsynonymous substitutions per nonsynonymous site per year and 1.6×10^{-9} synonymous substitutions per synonymous site per year, so that the overall rate is $3.5 \times 10^{-9} \pm 0.7$ per site per year; this is comparable to the rate of mammalian *Mhc* loci (Satta *et al*. 1993, Edwards *et al*. 1995). Taking the pheasant-fowl divergence at 20 MYA, we calibrated the scale in figure 3. The standard error of the pheasant-fowl distance estimate indicates the calibration error to be about 20%. The perusal of the scale indicates that all 21 alleles of Group 1 coalesce to a single common ancestor about 15 MYA. Since the formation of the Galápagos Archipelago did not begin until about 10 MYA (Christie *et al*. 1992; but see Geist 1996; furthermore there are data suggesting that marine iguanas of the Galápagos Archipelago diverged earlier than 10 MYA; see Rassmann 1995), the founding population of Darwin's finches presumably contained multiple allelic *Mhc* lineages, some of which have persisted up to the present time.

To estimate how many of the allelic lineages were present in the founding population, it is necessary to

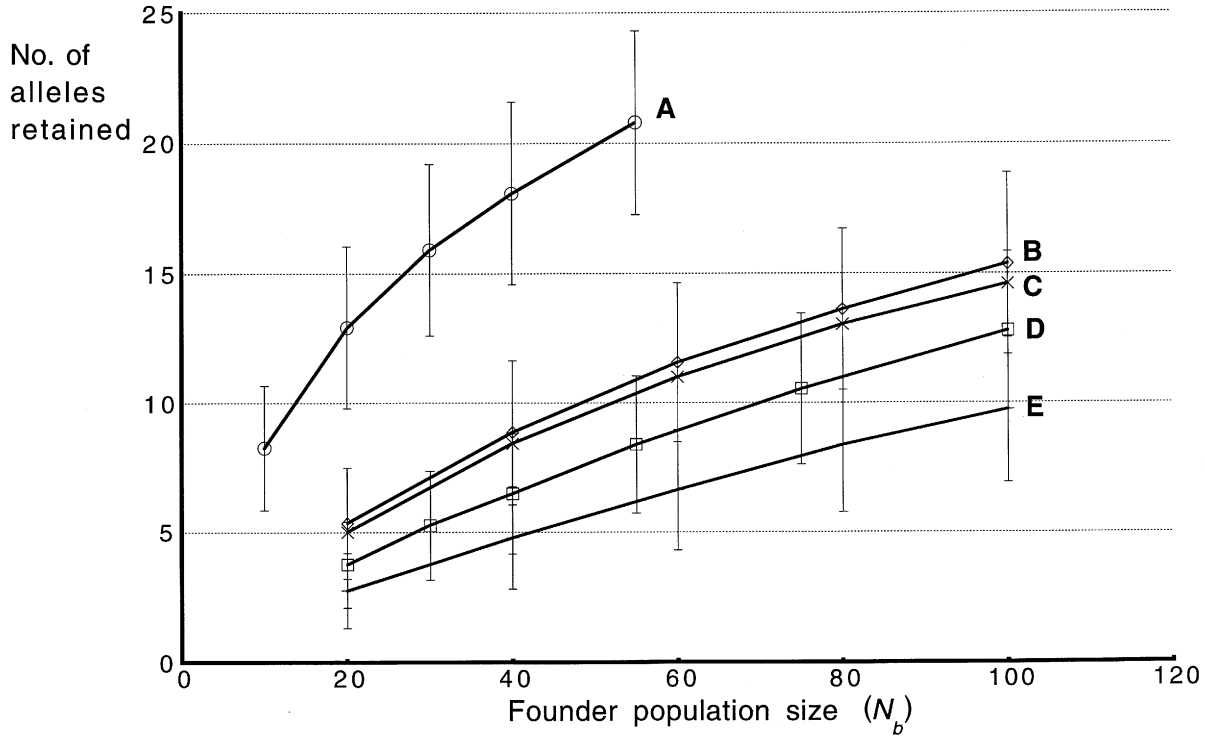

Figure 3. Computer simulation to determine the number of alleles retained in populations founded by individuals and evolved according to the program described in §2. Four models of evolution were tested: A, Immediate expansion from the founding population at a growth rate of 50 %; B, No expansion – population size remains constant at founding (N_b) level for ten generations; C, Expansion at a growth rate of 5 % begins after ten generations, during which the population size remained constant at the N_b level; D, The same as under A, but at a growth rate of 5 %. In all simulations s was equal to 0.01, and simulations A, C, and D were stopped once the population size of 1000 individuals was reached. The bars indicate 90 % confidence limits of the number of retained alleles (i.e. in 5 % of the cases the number of retained alleles was greater than indicated by the upper limit, and in another 5 % of cases it was lower than the indicated lower limit).

know when the founders arrived at the Islands. Obviously, they could not have arrived before the first island was formed, so the arrival time presumably was < 10 MYA. Two sets of data are available for narrowing down the arrival time more precisely – data on allozyme (Young & Patton 1981) and *Mhc* variation (this communication). The original estimate based on 27 allozyme-encoding loci suggested that the 13 species of Darwin's finches diverged 570000 years ago (Young & Patton 1981). Reanalyses of these same data using an updated calibration of the avian molecular clock puts the divergence time at 2.8 MYA (Grant 1994). The observations that identical *Mhc* sequences are shared by some of the finch species can be used to estimate the minimum divergence time of these species. Taking the estimated rate of 3.5×10^{-9} substitutions per site per year, it could be expected that the birds would accumulate one substitution per exon 2 per 0.6 million years (MY). The presence of identical alleles (in terms of exon 2 sequences) in different species therefore suggests that at least some of the species diverged < 0.6 MYA (the possibility cannot, however, be excluded that the sharing is the result of inter-species hybridization). Both the *Mhc*- and the allozyme-based estimates have a large range of error.

Since the arrival time of the founders at the Galápagos Islands cannot be determined precisely from the available data, we will consider three different values here: 600000 years (the lower limit estimated from the *Mhc* data), 2.8 MY (the revised estimate from the allozyme data), and 5 MY (the approximate age of the extant islands and the upper limit of all the estimates, see White *et al.* 1993). The initial colonization preceded the initial within-Archipelago divergence by an unknown period of time. From the dendogram in figure 2, we estimate that of the 21 alleles now constituting group one, 20, 16 and 13 allelic lineages existed 600000 years, 2.8 and 5 MYA, respectively. If all the birds that colonized the Islands were heterozygous for Group 1 alleles, and if no alleles were lost subsequent to colonization, the absolute minimum sizes of the founding population 0.6, 2.8 and 5 MYA would have to be 10, 8 and 7 individuals, respectively. To take the effect of drift into account and to estimate the number of founders needed to ensure the emergence of 21 Group 1 allelic lineages now present in the species tested, we carried out computer simulation using the program described in §2 and varying the input parameters.

The fate of the *Mhc* allelic lineages after the founding event could have been influenced by the mode of expansion of the founding population. We have, therefore, considered two variants of this parameter. In the first variant, the size of the population remained at the founding level for ten generations (= bottleneck period) and then began to expand logarithmically. In the second variant, the founding population began to expand immediately after the birds reached the islands.

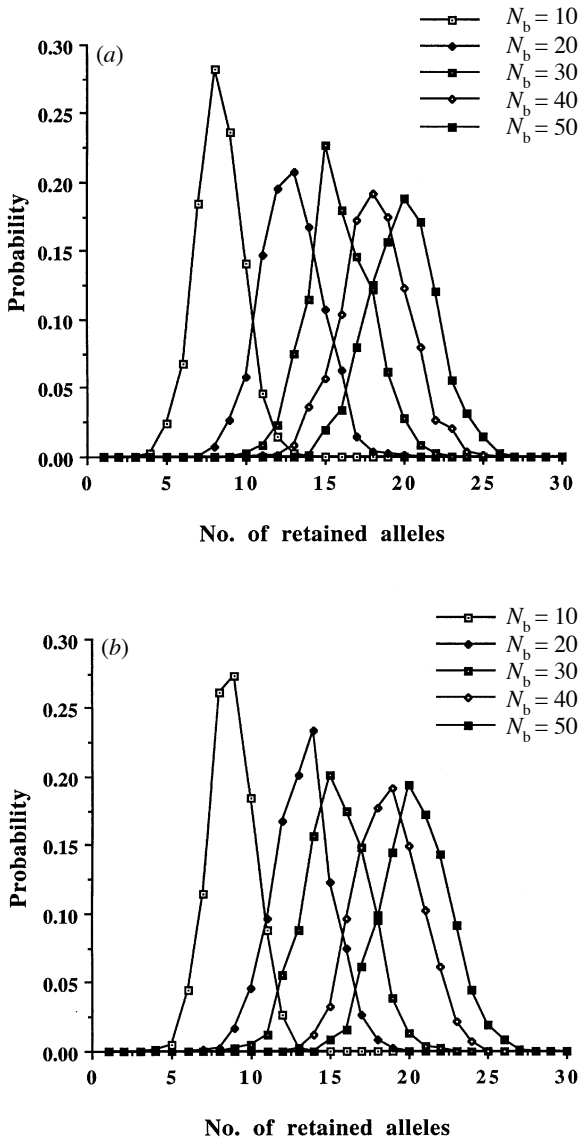

Figure 4. Probability distribution for the number of retained alleles based on model A computer simulation in figure 3. Note that because of the rapid growth rate and small N_b values, the two cases of (*a*) $s = 0$ and (*b*) $s = 0.5$ are almost indistinguishable. For $N_b = 20$ and $s = 0.5$, the probability of $k \geqslant 16$ and $\geqslant 20$ is 0.112 and 0, respectively, so that larger N_b values are required to retain those ancestral allelic lineages (where k is the number of retained alleles).

Current population sizes of most species on most islands fluctuate between hundreds and thousands of individuals (Grant & Grant 1989, 1992). We have examined population expansion up to a size of 1000 individuals in our models, after which we presume no allelic loss to occur. This assumption reduces the liberal estimates of founding population size somewhat, but conservative estimates are unaffected.

Parts of the *Mhc* exon 2 (the peptide-binding region or PBR-specifying sites) are known to be evolving under balancing selection pressure (Hughes & Nei 1989; Takahata *et al.* 1992). The intensity of this pressure (s) has been estimated as being < 0.01 (Satta *et al.* 1994). We have therefore taken this value of s into consideration.

In the simulations, we used three combinations of parameters that represent the extremes and a middle ground between them. The one extreme is: founding event taking place 5 MYA, expansion following founding immediately at a rate of 50%, and no further loss of alleles once the population exceeds 1000 individuals. This extreme minimizes the size of the founding population by providing the most favourable conditions for the persistence of alleles from the time of the founding event to the present. The other extreme is: founding 600000 years ago and bottleneck period lasting for ten generations followed by expansion to 1000 individuals at a rate of 5%. (We include this variant for completeness; in those instances in which founding of a new population has been observed, exponential growth occurred without a lag period; see Grant & Grant 1995.) These conditions maximize the size of the founding population. The third set of parameters (which we consider to be most realistic) is: founding event 2.8 MYA and expansion to 1000 individuals at a rate of 5% or 50% immediately following the founding event. The results of the simulation are summarized in figures 3 and 4. The simulation shows that the most favourable conditions for the retention of alleles exist when the population begins to expand at a high rate of 50% (model A in figure 3) immediately after founding. These conditions therefore provide the most conservative estimate of N_b; all other conditions tested (figure 3B–E) require higher N_b values than model A for retention of alleles. The population of $N_b = 50$ (40) under model A will retain 20 or more alleles with a probability of 67% (34%); populations of smaller founding sizes are less likely to do so, the probability being 2% for $N_b = 30$ and 0% for $N_b = 20$. The probability of retaining no fewer than 16 or 13 alleles is higher: 99% or 100% for $N_b = 50$%, 95% or 100% for $N_b = 40$, 48% or 93% for $N_b = 30$, and 11% or 67% for $N_b = 20$. A founder population of 10 individuals cannot retain even 13 alleles (figure 4). In other words, to retain at least 20, 16 and 13 alleles with a probability of 5%, the founding population sizes would have to be > 32, > 20 and > 15 individuals, respectively (figure 3). Hence, even if the colonization of the Galápagos Islands occurred as long as 5 MYA, founding by fewer than 15 birds would be highly improbable.

These estimates are most conservative not only because model A is the most favourable for the retention of alleles, but also for other reasons: the number of Group 1 alleles is almost certainly greater than 21 (the tested sample was small and not all the species and islands were sampled); only one of several *Mhc* loci was used for the calculations (inclusion of other loci would have increased the N_b estimates); the use of the Caucasoid population as a model for the parental finch population probably overestimates the variability of the latter (and hence underestimates the N_b); and the assumption of no allelic loss after the population reached the size of 1000 individuals is probably unrealistic. The factors that could reduce the N_b estimate include the time since colonization, intensity of selection, and the *Mhc* substitution rate. As was indicated earlier, it is not likely that the Islands were colonized more than 5 MYA. As for selection

intensity, computer simulations have revealed very little influence of the various s values on the N_b estimates (figure 4), probably because in the relatively small populations drift prevails over selection. In particular, the initial sampling of N_e genes strongly influences the number and frequencies of founding alleles. Similarly, variation in the substitution rate within the standard error of the estimate has only a minor effect on the population size estimate.

While the evidence for the evolutionary rate at *Mhc* loci used in the simulation is strong (see Klein *et al.* 1993*b*), one might perhaps argue that under exceptional circumstances such as those existing when a population colonizes an entirely new environment, the rate might be accelerated. It has indeed been claimed, for example, that human *Mhc* class I loci evolve rapidly in certain South American Indian populations because of selection pressure exerted by the environment (Belich *et al.* 1992; Watkins *et al.* 1992). We have, therefore, considered the possibility that most of the *Mhc* polymorphism now found in the Darwin's finches arose by mutation and has been maintained by selection after the colonization of the Galápagos Islands. There is convincing evidence that positive (balancing) selection influences the evolution of functional *Mhc* loci (Hughes & Nei 1989; Takahata *et al.* 1992) and that the effect is limited to the sites encoding the PBR of the *Mhc* molecule. These sites are known in humans and mice (Brown *et al.* 1993) and can be identified in other species by their polymorphism. We therefore eliminated the putative PBR sites from the sequence alignments and reanalysed the sequences without them. This modification, however, did not influence significantly the results of the analysis. Even when the PBR sites are not taken into account, it is still necessary to postulate the presence, 2.8 MYA, in the founding population of at least 14 alleles at the Group 1 locus to explain the extant polymorphism at this locus.

If the effective population size (N_e) was 30 individuals the number of actual colonists (N) is likely to have been 50–100 because effective sizes are 25–50% of the number of breeders (Grant & Grant 1992). Darwin's finches form large flocks of this size and disperse in the non-breeding season (Grant 1986), and it is reasonable to assume that their ancestors did so as well. Flocks are largest, and inter-island movements are most frequent, during and after an El Niño year of prolific breeding, such as occurred in 1983. Vegetation mats get carried out to sea from the estuary of the Guayas river on the adjacent mainland, especially in El Niño years, and one of these may have acted as a resting place for the colonizing finches at some intermediate distance between the mainland and the islands. Indeed, the occurrence on the Galápagos of non-flying and non-swimming terrestrial animals like land iguanas, lizards and geckos may best be explained by rafting across the whole distance. While these ideas are necessarily speculative, long-distance colonizations by birds have been recorded elsewhere in modern times. White-eyes (*Zosterops lateralis*) colonized the south island of New Zealand from Tasmania in 1856, having crossed a gap of 2000 km (Mayr 1963), and fieldfares (*Turdus pilaris*) colonized Greenland, apparently from southern Norway and via Jan Mayen Island, in 1937 (Salomonsen 1951). In neither case was the size of the founding population known, although the fieldfares were known to have travelled in a large flock (Salomonsen 1951). Another possibility is that the islands were colonized more than once by a smaller number of individuals each time. The remoteness of the islands make this unlikely, although not impossible. All we can say is that there is no clear sign of multiple colonizations from different points of origin in the distribution of the *Mhc* polymorphism among the extant species. We conclude, therefore, that the Galápagos Islands were colonized once by the ancestors of Darwin's finches, and that the founding population consisted of no fewer than 30 individuals. We emphasize that this is the minimum estimate obtained by considering extreme conditions favouring colonization; it is, in all probability, an underestimate of the actual size of the founding population.

The main implication of our estimate is that a major genetic reorganization associated with an extreme bottleneck (e.g. see Templeton 1980; Carson & Templeton 1984) is unlikely to have occurred in the early stages of establishment of Darwin's finches on the Galápagos Islands.

We thank the Galápagos National Park Service and the Charles Darwin Research Station for permission and support in carrying out the field work, Ms Moira Burghoffer for editorial assistance, Mr Joseph Olczyk for technical assistance, and Dr David Stern, Department of Ecology and Evolutionary Biology, Princeton University, for critical reading of the manuscript.

REFERENCES

Belich, M. P., Madrigal, J. A., Hildebrand, W. H., Zemmour, J., Williams, R. C., Luz, R., Petzl-Erler, M. L. & Parham, P. 1992 Unusual *HLA-B* alleles in two tribes of Brazilian Indians. *Nature, Lond.* **357**, 326–329.

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L. & Wiley, D. C. 1993 Three-dimensional structure of the human class II histocompatibility antigen HLA-DR. *Nature, Lond.* **364**, 33–39.

Carson, H. L. & Templeton, A. R. 1984 Genetic revolutions in relation to speciation phenomena: the founding of new populations. *Annu. Rev. Ecol. Syst.* **15**, 97–131.

Christie, D. M., Duncan, R. A., McBirney, A. R., Richards, M. A., White, W. M., Harpp, K. S. & Fox, C. B. 1992 Drowned islands downstream from the Galápagos hotspot imply extended speciation times. *Nature, Lond.* **355**, 246–248.

Cox, A. 1983 Ages of the Galápagos Islands. In *Patterns of evolution in Galápagos organisms* (ed. R. I. Bowman, M. Berson, & A. E. Leviton), pp. 11–23. San Francisco, CA: Amer. Assoc. Adv. Sci., Pacific Division.

Coyne, J. A., Crittenden, A. P. & Mah, K. 1994 Genetics of a pheromonal difference contributing to reproductive isolation in *Drosophila. Science, Wash.* **265**, 1461–1464.

Dobzhansky, T. 1937 *Genetics and the origin of species*. New York: Columbia University Press.

Edwards, S. V., Wakeland, E. K. & Potts, W. K. 1995 Contrasting histories of avian and mammalian *Mhc* genes revealed by class II B sequences from songbirds. *Proc. natn. Acad. Sci. U.S.A.* **92**, 12200–12204.

Geist, D. 1996 On the emergence and submergence of the Galápagos Islands. *Noticias de Galápagos* **56**, 5–9.

Giddings, L. V., Kaneshiro, K. Y. & Anderson, W. W. 1989 (eds) *Genetics, speciation, and the founder principle*. New York: Oxford University Press.

Grant, P. R. 1986 *Ecology and evolution of Darwin's finches*. Princeton University Press.

Grant, P. R. 1994 Population variation and hybridization: comparison of finches from two archipelagos. *Evol. Ecol.* **8**, 598–617.

Grant, B. R. & Grant, P. R. 1989 *Evolutionary dynamics of a natural population. The large cactus finch of the Galápagos*. Chicago: University Press.

Grant, P. R. & Grant, B. R. 1992 Demography and the genetically effective sizes of two populations of Darwin's finches. *Ecology* **73**, 766–784.

Grant, P. R. & Grant, B. R. 1995 The founding of a new population of Darwin's finches. *Evolution* **49**, 229–240.

Helm-Bychowski, K. M. & Wilson, A. C. 1986 Rates of nuclear DNA evolution in pheasant-like birds: evidence from restriction maps. *Proc. natn. Acad. Sci. U.S.A.* **83**, 688–692.

Hollocher, H. 1996 Island hopping in *Drosophila*: patterns and processes. *Phil. Trans. R. Soc. Lond.* B **351**, 735–743.

Hudson, R. R. 1990 Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* **7**, 1–44.

Hughes, A. L. & Nei, M. 1989 Nucleotide substitution at major histocompatibility complex class II loci: Evidence for overdominant selection. *Proc. natn. Acad. Sci. U.S.A.* **86**, 958–962.

Imanishi, T., Wakisaka, A., & Gojobori, T. 1991 Genetic relationships among various human populations indicated by MHC polymorphisms. In *HLA 1991. Proceedings of the Eleventh International Histocompatibility Workshop and Conference, Yokohama, Japan*, vol 1. (ed. K. Tsuji, M. Aizawa & T. Sasazuki), pp. 627–632. Oxford: Oxford University Press.

Jackson, M. H. 1993 *Galápagos. A natural history*. Calgary: University of Calgary Press.

Kimura, M. 1980 A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120.

Kingman, J. F. C. 1982*a* The coalescent. *Stochastic Proc. Appl.* **13**, 235–248.

Kingman, J. F. C. 1982*b* On the genealogy of large populations. *J. Appl. Prob.* **13**, 27–43.

Klein, J. 1986 *Natural history of the major histocompatibility complex*. New York: Wiley.

Klein, J. 1987 Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. *Hum. Immunol.* **19**, 155–162.

Klein, J. & Figueroa F. 1986 Evolution of the major histocompatibility complex. *CRC Crit. Rev. Immunol.* **6**, 295–386.

Klein, J., Gutknecht, J. & Fischer, N. 1990*a* The major histocompatibility complex and human evolution. *Trends Genet.* **6**, 7–11.

Klein, J., Kasahara, M., Gutknecht, J., & Figueroa, F. 1990*b* Origin and function of Mhc polymorphism. *Chemical Immunology* **49**, 35–50.

Klein, J., Bontrop, R. E., Dawkins, R. L *et al.* 1990*c* Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics* **31**, 217–219.

Klein, D., Ono, H., O'Huigin, C., Vincek, V., Goldschmidt, T. & Klein, J. 1993 Extensive MHC variability in cichlid fishes of Lake Malawi. *Nature, Lond.* **364**, 330–334.

Klein, J., Satta, Y., Takahata, N. & O'Huigin, C. 1993*a* Trans-specific *Mhc* polymorphism and the origin of species in primates. *J. Med. Primatol.* **22**, 57–64.

Klein, J., Satta, Y., O'Huigin, C. & Takahata, N. 1993*b* The molecular descent of the major histocompatibility complex. *A. Rev. Immunol.* **11**, 269–295.

Lack, D. 1947 *Darwin's finches*. Cambridge University Press.

Mayr, E. 1942 *Systematics and the origin of species*. New York: Columbia University Press.

Mayr, E. 1963 *Animal species and evolution*. Cambridge, MA: Harvard University Press.

Nei, M. 1987 *Molecular evolutionary genetics*. New York: Columbia University Press.

Ono, H., O'Huigin, C., Tichy, H. & Klein, J. 1993 Major histocompatibility complex-variation in two species of cichlid fishes from Lake Malawi. *Molec. Biol. Evol.* **10**, 1060–1072.

Otte, D. & Endler, J. A. (eds) 1989 *Speciation and Its Consequences*. Sunderland, MA: Sinauer Associates.

Rassmann, K. 1995 Das Atlantis der Leguane. Wie alt ist Galápagos? *Praxis der Naturwissenschaften – Biologie* **44** (8), 5–9.

Rice, W. R. & Hostert, E. E. 1993 Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*, **47**, 1637–1653.

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R. & Horn, G. T. 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science, Wash.* **239**, 487–491.

Saitou, N. & Nei, M. 1987 The neighbor-joining method. A new method for reconstructing phylogenetic trees. *Molec. Biol. Evol.* **4**, 406–425.

Salomonsen, F. 1951 The immigration and breeding of the fieldfare (*Turdus pilaris* L.) in Greenland. *Proc. Xth Int. Ornithol. Congr. Uppsala, 1950*, pp. 515–526.

Sanger, F., Nicklen, S. & Coulson, A. R. 1977 DNA sequencing with chain-terminating inhibitors. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467.

Satta, Y., O'Huigin, C., Takahata, N. & Klein, J. 1993 The synonymous substitution rate of the major histocompatibility complex loci in primates. *Proc. natn. Acad. Sci. U.S.A.* **90**, 7480–7484.

Satta, Y., O'Huigin, C., Takahata, N. & Klein, J. 1994 Intensity of natural selection at the major histocompatibility complex loci. *Proc. natn. Acad. Sci. U.S.A.* **91**, 7184–7188.

Simkin, T. 1984 Geology of Galápagos. *Biol. J. Linn. Soc.* **21**, 61–65.

Sokal, R. R. & Michener, C. D. 1958 A statistical method for evaluating systematic relationships. *University of Kansas Sci. Bull.* **28**, 1409–1438.

Srivastava, R., Ram, B. P., & Tyle, P. 1991 *Immunogenetics of the major histocompatibility complex*. New York: VCH Publishers.

Takahata, N. 1990 A simple genealogical structure of strongly balanced allelic lines and trans-species evolution of polymorphism. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2419–2423.

Takahata, N. & Nei, M. 1990 Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. *Genetics* **124**, 967–978.

Takahata, N., Satta, Y. & Klein, J. 1992 Polymorphism and balancing selection at major histocompatibility complex. *Genetics* **130**, 925–938.

Tavaré, S. 1984 Line-of-descent and genealogical processes, and their applications in population genetics models. *Theor. Popul. Biol.* **26**, 119–164.

Templeton, A. R. 1980 The theory of speciation via the founder principle. *Genetics* **94**, 1011–1038.

Templeton, A. R. 1981 Mechanisms of speciation – a population genetic approach. *A. Rev. Ecol. Syst.* **12**, 23–48.

Vincek, V., Klein, D., Graser, R. T., Figueroa, F., O'Huigin, C. & Klein, J. 1995 Molecular cloning of major histocompatibility complex class II *B* gene from the Bengalese finch *Lonchura striata*. *Immunogenetics* **42**, 262–267.

Watkins, D. I., McAdam, S. N., Liu, X. *et al*. 1992 New recombinant HLA-B alleles in a tribe of South Amerindians indicate rapid evolution of MHC class I loci. *Nature*, *Lond*. **357**, 329–333.

Watterson, G. A. 1984 Lines of descent and the coalescent. *Theor*. *Popul*. *Biol*. **26**, 77–92.

White, W. M., McBirney, A. R. & Duncan, R. A. 1993 Petrology and geochemistry of the Galápagos Islands: portrait of a pathological mantle plume. *J*. *geophys*. *Res*. **98**, 19533–19563.

Wittzell, H., von Schantz, T., Zoorob, R. & Auffray, C. 1994 Molecular characterization of three *Mhc* class II *B* haplotypes in the ring-necked pheasant. *Immunogenetics* **39**, 395–403.

Wu, C.-I., Hollocher, H., Begun, D. J., Aquandro, C. F., Xu, Y. & Wu, M.-I. 1995 Sexual isolation in *Drosophila melanogaster*: a possible case of incipient speciation. *Proc*. *natn*. *Acad*. *Sci*. *U.S.A*. **92**, 2519–2523.

Young, S. Y. & Patton, J. L. 1981 Genetic variability and differentiation in Galápagos finches. *The Auk* **98**, 230–242.

Zoorob, R., Behar, G., Kroemer, G. & Auffray, C. 1990 Organization of a functional chicken class II *B* gene. *Immunogenetics* **31**, 179–187.

Received 15 August 1996; accepted 9 September 1996.