The ages for MJG-I, MJG-II and MJG-III are considerably older than previous age estimates of Palaeolithic sites in northern China[1] and indicate that humans might have reached northeast Asia earlier than previously thought. Along with estimated ages for the sites of Gongwangling (1.15 Myr)[14] and Xihoudu (1.27 Myr)[15] in the southern Loess Plateau and for Xiaochangliang (1.36 Myr)[1] and Donggutuo (1.1 Myr)[16] sites in the Nihewan basin, our new results imply an expansion and lengthy flourishing of human groups from northern to north-central China during the early Pleistocene.

The estimated age of 1.66 Myr for the MJG-III artefact layer predates the previous oldest age of unambiguous human presence at 40° N in East Asia by about 0.3 Myr. Our findings, particularly for the MJG-III layer, document the oldest coexistence of stone tools and man-made bone modifications in East Asia, indicating possible continuity with the oldest stone tools and artificial bone modifications reported in eastern Africa[17,18]. Archaeological evidence at MJG indicates the oldest known use of animal tissues, especially marrow processing, by early humans in Asia. The earliest archaeological level in the Nihewan basin is slightly younger than the 1.75 Myr estimated age for early humans at the Dmanisi site at 40° N latitude in western Eurasia[2,3]. Our estimated ages also fall within the 1.66–1.51-Myr range for the earliest known human fossils in southeast Asia[19,20]. The combined evidence suggests that, near the start of the Pleistocene, early human populations spread relatively rapidly across Asia, presumably from an African origin, and reached at least 40° N latitude. Our findings further establish that the earliest populations to reach northeast Asia were able to survive for at least 500 kyr before the mid-Pleistocene onset of high-amplitude climate oscillation[21–23]. □

Received 19 February; accepted 8 July 2004; doi:10.1038/nature02829.

1. Zhu, R. X. *et al.* Earliest presence of humans in northeast Asia. *Nature* **413,** 413–417 (2001).
2. Gabunia, L. *et al.* Earliest Pleistocene hominid cranial remains from Dmanisi, Republic of Georgia: Taxonomy, geological setting, and age. *Science* **288,** 1019–1025 (2000).
3. Vekua, A. *et al.* A new skull of early *Homo* from Dmanisi, Georgia. *Science* **297,** 85–89 (2002).
4. Wei, Q. Banshan Paleolithic site from the lower Pleistocene in the Nihewan Basin in northern China. *Acta Anthropol. Sinica* **13,** 223–238 (1994).
5. HPICR, *Papers on Archaeology in Hebei Province* 30–45 (East, Beijing, 1998).
6. Potts, R. *Early Hominid Activities at Olduvai* (Aldine de Gruyter, New York, 1988).
7. Potts, R., Behrensmeyer, A. K. & Ditchfield, P. Paleolandscape variation and Early Pleistocene hominid activities: Members 1 and 7, Olorgesailie Formation, Kenya. *J. Hum. Evol.* **37,** 747–788 (1999).
8. Tang, Y. J., Li, Y. & Chen, W. Y. Mammalian fossils and the age of Xiaochangliang paleolithic site of Yangyuan, Hebei. *Vertebrata Palasiatica* **33,** 74–83 (1995).
9. Berggren, W. A., *et al.* in *Geochronology, Timescales, and Stratigraphic Correlation* (eds Berggren, W. A., Kent, D. V., Aubry, M. & Hardenbol, J.) 129–212 (SEPM Spec. Publ. 54, Tulsa, Oklahoma, 1995).
10. Wei, Q., *et al.* in *Evidence for Evolution—Essays in Honor of Prof. Chungchien Yong on the Hundredth Anniversary of His Birth* (ed. Tong, Y.) 193–207 (Ocean, Beijing, 1997).
11. Huang, W. P. & Fang, Q. R. *Wushan Hominid Site* 105–109 (Ocean, Beijing, 1991).
12. Qiu, Z. X. Nihewan fauna and Q/N boundary in China. *Quat. Sci.* **20,** 142–154 (2000).
13. Singer, B. S. *et al.* Dating transitionally magnetized lavas of the late Matuyama chron: Toward a new $^{40}Ar/^{39}Ar$ timescale of reversals and events. *J. Geophys. Res.* **104,** 679–693 (1999).
14. An, Z. S. & Ho, C. K. New magnetostratigraphic dates of Lantian *Homo erectus*. *Quat. Res.* **32,** 213–221 (1989).
15. Zhu, R., An, Z., Potts, R. & Hoffman, K. A. Magnetostratigraphic dating of early humans in China. *Earth Sci. Rev.* **61,** 341–359 (2003).
16. Quaternary Research Association of China, Li, H. M. & Wang, J. D. *Quaternary Geology and Environment of China* 33–38 (Ocean, Beijing, 1982).
17. Semaw, S. *et al.* 2.5-million-year-old stone tools from Gona, Ethiopia. *Nature* **385,** 333–336 (1995).
18. de Heinzelin, J. *et al.* Environment and behavior of 2.5-million-year-old Bouri hominids. *Science* **284,** 625–629 (1999).
19. Swisher, C. C. III *et al.* Age of the earliest known hominids in Java, Indonesia. *Science* **263,** 1118–1121 (1994).
20. Larick, R. *et al.* Early Pleistocene $^{40}Ar/^{39}Ar$ ages for Bapang Formation hominins, Central Java, Indonesia. *Proc. Natl Acad. Sci. USA* **98,** 4866–4871 (2001).
21. Potts, R. in *Human Roots: Africa and Asia in the Middle Pleistocene* (eds Barham, L. & Robson-Brown, K.) 5–21 (Western Academic & Specialist Press, Bristol, 2001).
22. Clark, P. U., Alley, R. B. & Pollard, D. Northern Hemisphere ice-sheet influences on global climate change. *Science* **286,** 1104–1111 (1999).
23. Tian, J., Wang, P., Cheng, X. & Li, Q. Astronomically tuned Plio-Pleistocene benthic $\delta^{18}O$ record from South China Sea and Atlantic–Pacific comparison. *Earth Planet. Sci. Lett.* **203,** 1015–1029 (2002).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank R. J. Enkin for providing palaeomagnetic software. This work was supported by the National Natural Science Foundation of China and Chinese Academy of Sciences. R.P. was supported by the US National Science Foundation and the Smithsonian Human Origins Program. K.A.H. also received support from the US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.X.Z. (rxzhucn@yahoo.com and rxzhu@mail.igcas.ac.cn) or R.P. (potts.rick@nmnh.si.edu).

Modelling the recent common ancestry of all living humans

Douglas L. T. Rohde[1], Steve Olson[2] & Joseph T. Chang[3]

[1]*Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*7609 Sebago Road, Bethesda, Maryland 20817, USA*
[3]*Department of Statistics, Yale University, New Haven, Connecticut 06520, USA*

If a common ancestor of all living humans is defined as an individual who is a genealogical ancestor of all present-day people, the most recent common ancestor (MRCA) for a randomly mating population would have lived in the very recent past[1–3]. However, the random mating model ignores essential aspects of population substructure, such as the tendency of individuals to choose mates from the same social group, and the relative isolation of geographically separated groups. Here we show that recent common ancestors also emerge from two models incorporating substantial population substructure. One model, designed for simplicity and theoretical insight, yields explicit mathematical results through a probabilistic analysis. A more elaborate second model, designed to capture historical population dynamics in a more realistic way, is analysed computationally through Monte Carlo simulations. These analyses suggest that the genealogies of all living humans overlap in remarkable ways in the recent past. In particular, the MRCA of all present-day humans lived just a few thousand years ago in these models. Moreover, among all individuals living more than just a few thousand years earlier than the MRCA, each present-day human has exactly the same set of genealogical ancestors.

In investigations of the common ancestors of all living humans, much attention has focused on descent through either exclusively maternal or exclusively paternal lines, as occurs with mitochondrial DNA and most of the Y chromosome[4,5]. But according to the more common genealogical usage of the term 'ancestor', ancestry encompasses all lines of descent through both males and females, so that the ancestors of an individual include all of that person's parents, grandparents, and so on.

For a population of size n, assuming random mating (and so ignoring population substructure), probabilistic analysis[2] has proved that the number of generations back to the MRCA, T_n, has a distribution that is sharply concentrated around $\log_2 n$. We express this using the notation $T_n \sim \log_2 n$, meaning that the quotient $T_n/\log_2 n$ converges in probability to 1 as $n \to \infty$. In contrast, the mean time to the MRCA along exclusively matrilineal or patrilineal lines is approximately n generations[6], and the distribution is not sharply concentrated. For example, in a panmictic population of one million people, the genealogical MRCA would have lived about 20 generations ago, or around the year AD 1400, assuming a generation time of 30 years. The MRCA along

Box 1
Graph-theoretical definitions

The length of a path in a graph, G, is the number of edges in the path. For each pair of nodes i and j in G, the distance $d(i, j)$ is defined to be the length of a shortest path joining i and j. The radius of G is

$$R = \min_{i \in G}\{\max_{k \in G} d(i,k)\}$$

and a node i is called a centre of G if $\max_{k \in G} d(i, k) = R$. Assume $R \geq 1$; the case $R = 0$ (G has one node) was treated previously[2]. For each centre node i, let S_i be a set of minimal size that consists of neighbours of node i and satisfies $\min\{d(j, k): j \in \{i\} \cup S_i\} \leq R - 1$ for all $k \in G$. H_i is defined as the number of nodes in S_i, H is the minimum of H_i over all centres i, and $\Delta = 1 - \frac{1}{H}$. The diameter of G is $D = \max_{i,k \in G} d(i, k)$.

exclusively maternal lines would have lived something like 50,000 times earlier—in the order of one million generations ago.

As genealogical ancestry is traced back beyond the MRCA, a growing percentage of people in earlier generations are revealed to be common ancestors of the present-day population. Tracing further back in time, there was a threshold, let us say U_n generations ago, before which ancestry of the present-day population was an all or nothing affair. That is, each individual living at least U_n generations ago was either a common ancestor of all of today's humans or an ancestor of no human alive today. Thus, among all individuals living at least U_n generations ago, each present-day human has exactly the same set of ancestors. We refer to this point in time as the identical ancestors (IA) point. As with the MRCA point, the IA point is also quite recent in a randomly mating population: $U_n \sim 1.77 \log_2 n$ generations ago[2].

The major problem in applying these results to human populations is that mating is not random in the real world. Mating patterns are structured by geography, proximity, culture, language and social class. Nevertheless, even in populations with considerable internal structure, the time to the MRCA can be remarkably brief. To demonstrate this in a tractable mathematical model, consider a population of size n divided into randomly mating subpopulations that are linked by occasional migrants. The population is represented by a graph, G, with a node for each subpopulation. Edges indicate pairs of nodes that exchange a small number (for example, one pair) of migrants per generation. Let R denote the radius of G, and let Δ be a quantity ranging between 0 and 1 that depends on the structure of G (see Box 1). A probabilistic analysis (see Supplementary Information) shows that as $n \to \infty$, $T_n \sim (R + \Delta)\log_2 n$. Furthermore, if we let D denote the diameter of the graph, then the number of generations, U_n, since the IA point satisfies $U_n \sim (D + 1.77)\log_2 n$.

Computer simulations accord with these theoretical predictions. Tables 1 and 2 give distributions of T_n and U_n for small populations of varying sizes in graphs with one node, three connected nodes, five fully connected nodes and for a ten-node graph loosely based on world geography as shown in Fig. 1. In these simulations, neighbouring subpopulations exchange one pair of migrants per generation. Each mean is calculated from 100 model runs. Although guaranteed to be accurate only for sufficiently large n, the theoretical predictions describe the simulations quite well even for models with just a few thousand individuals. Whenever n is doubled, T_n is expected to increase by $R + \Delta$, and U_n is expected to increase by $D + 1.77$. These predicted increases, which are listed in the last columns of Tables 1 and 2, agree closely with the simulation results.

To hazard a rough first guess about human recent common ancestors, we could extrapolate the results for the graph of Fig. 1 to a growing population with a final size of 250 million. When applying this model to a growing population, the fixed population size that provides the best approximation is the size at the time that the MRCA lived. We take this effective population size to be 250 million, which is approximately the global population in the year AD 1. Starting from $n = 16{,}000$, a population of 250 million is reached by doubling 14 times. Approximating the increases in T_n and U_n beyond the values seen in Tables 1 and 2 by their theoretical predictions for each doubling of n, we arrive at $T_n \approx 34 + 14 \times 3 = 76$ generations (about 2,300 years) and $U_n \approx 74 + 14 \times 6.77 = 169$ generations (about 5,000 years). These estimates would suggest, with the exchange of just one pair of migrants per generation between large panmictic populations of realistic size, that the MRCA appears in about the year 300 BC, and all modern individuals have identical ancestors by about 3,000 BC. Such estimates are extremely tentative, and the model contains several obvious sources of error, as it was motivated more by considerations of theoretical insight and tractability than by realism. Its main message is that substantial forms of population subdivision can still be compatible with very recent common ancestors.

The dynamics of human subpopulations are much more complex than those in the simple graph model discussed above. Although these complexities make theoretical analysis difficult, a computer model incorporating more complicated forms of population substructure and migration allows the demographic history of human populations to be simulated. The Supplementary Information contains more details on the model and computations; here we briefly outline some of the main points.

This model is based on a simplified projection of the world's

Table 1 **Simulations of T_n**

Graph	$n = 1{,}000$	$n = 2{,}000$	$n = 4{,}000$	$n = 8{,}000$	$n = 16{,}000$	$R + \Delta$
One node	10.8 (0.4)	11.8 (0.4)	12.8 (0.4)	13.9 (0.3)	14.8 (0.4)	1.00
Three fully connected nodes	14.0 (0.7)	15.6 (0.7)	17.1 (0.9)	18.9 (0.8)	20.3 (1.0)	1.50
Five fully connected nodes	14.0 (0.5)	15.8 (0.5)	17.8 (0.5)	19.6 (0.5)	21.5 (0.6)	1.75
Ten-node graph shown in Fig. 1	21.1 (1.3)	24.3 (1.5)	27.6 (1.5)	30.5 (1.5)	33.8 (1.7)	3.00

Means (standard deviations in parentheses) of T_n (the number of generations back to the MRCA) for graph-structured populations exchanging a single pair of migrants per edge per generation. The last column shows $R + \Delta$, the expected asymptotic increase in T_n per doubling of n.

Table 2 **Simulations of U_n**

Graph	$n = 1{,}000$	$n = 2{,}000$	$n = 4{,}000$	$n = 8{,}000$	$n = 16{,}000$	$D + 1.77$
One node	20.8 (1.6)	22.6 (1.5)	24.6 (1.5)	26.5 (1.6)	28.3 (1.4)	1.77
Three fully connected nodes	27.4 (1.5)	30.3 (1.4)	33.4 (1.5)	36.2 (1.7)	38.9 (1.5)	2.77
Five fully connected nodes	25.9 (1.3)	28.9 (1.4)	32.1 (1.7)	35.3 (1.5)	37.9 (1.4)	2.77
Ten-node graph shown in Fig. 1	46.3 (2.7)	53.0 (2.7)	59.8 (2.7)	66.8 (2.9)	73.6 (2.7)	6.77

Means (standard deviations in parentheses) of U_n (the number of generations back to the IA point) for graph-structured populations exchanging a single pair of migrants per edge per generation. The last column shows $D + 1.77$, the expected asymptotic increase in U_n per doubling of n.

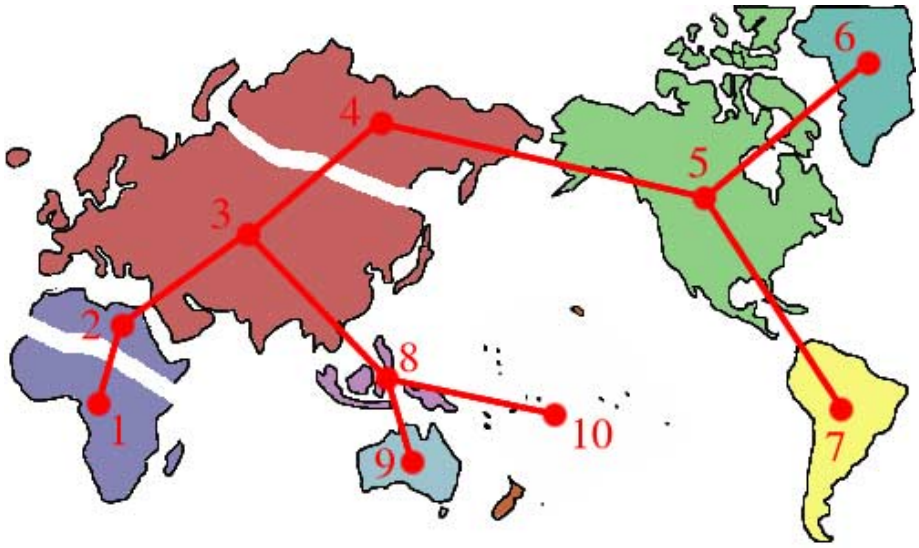

Figure 1 World map viewed as a ten-node graph. This graph has radius 3 and diameter 5.

actual inhabited land masses and has three levels of substructure: continents, 'countries' and 'towns.' Figure 2 depicts the model's geography and migration routes used before AD 1500, with the countries shown as squares and the number of towns per country differing from continent to continent. Towns and countries represent both the local geographical areas and the relevant social and ethnic groups from which most people find mates.

The model uses a simplified migration system in which each person has a single opportunity to migrate from his or her town of birth. The probabilities of leaving a town or a country are set at various levels to reflect different migration patterns. Migrants who move between towns can travel to any other town within the country. A migrant who leaves a country for another country within the same continent chooses the destination with a probability that diminishes as the inverse square of the geographical distance.

Each continent has a number of port countries from which migrants can travel to another continent. A fixed, large percentage (for example, 95% in some simulations) of the migrants through a port come from the country in which the port is located, with the remainder drawn from other countries in the continent in proportion to their inverse squared distance. The value next to a port in Fig. 2 is its migration rate, in people per generation, and the date in parentheses indicates when the port opens, if it is more recent than the start of the simulation in 20,000 BC. When a port opens, there is usually a single generation of migration at a higher rate than the steady-state rate shown in the figure. After the year AD 1500, additional large ports, which are not shown, begin to open to simulate colonization of the Americas, Australia and elsewhere. Immediately before this, the native population of the Americas is markedly reduced to simulate the effects of European-introduced diseases[7].

Generations overlap in this model and we explicitly simulated the lifespan and the times at which mating and reproduction events occur for each individual[8,9], as described in more detail in Supplementary Information. The birth rate of each continent or island was individually adjusted so that the populations match historical

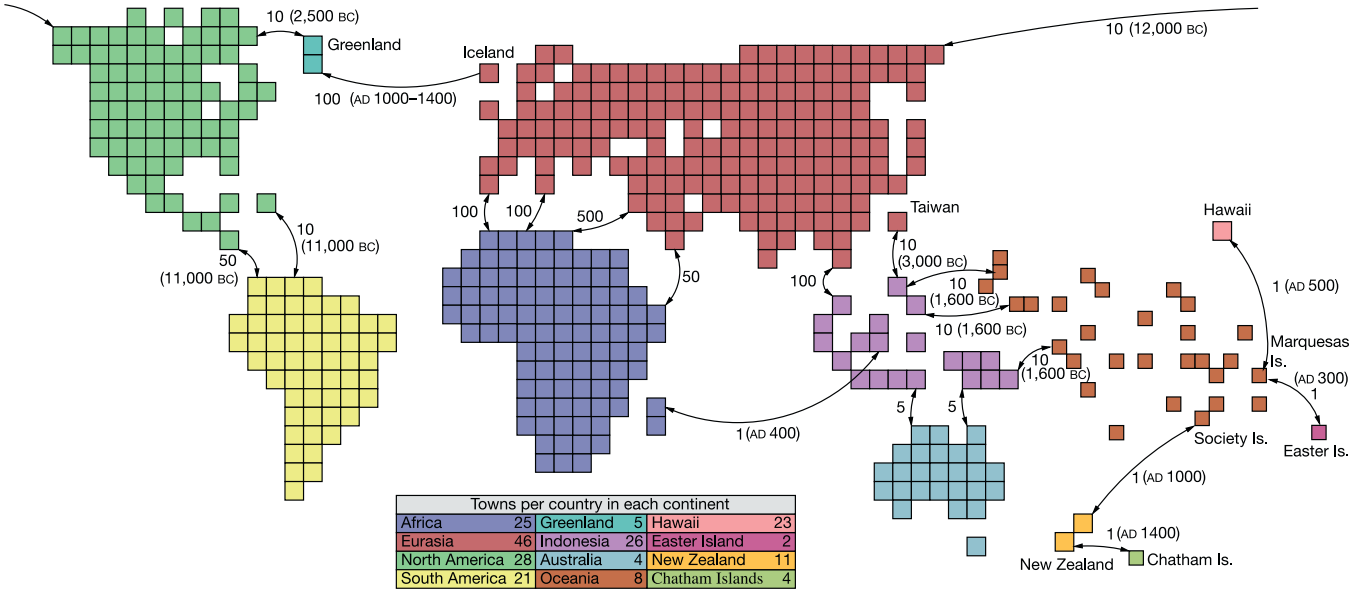

Figure 2 Geography and migration routes of the simulated model. Arrows denote ports and the adjacent numbers are their steady migration rates, in individuals per generation. If given, the date in parentheses indicates when the port opens. Upon opening, there is usually a first-wave migration burst at a higher rate, lasting one generation.

estimates, and growth rates were higher in under-populated areas. Full-sized populations were used until the world population reached 50 million in 1,000 BC. Subsequently, birth rates were reduced to achieve a worldwide level of 55 million, carried out in such a way that sparsely populated areas were less affected. This limit was a computational necessity, but simulations show that population growth has little effect, especially if it occurs after the MRCA has died.

With 5% of individuals migrating out of their home town, 0.05% migrating out of their home country, and 95% of port users born in the country from which the port emanates, the simulations produce a mean MRCA date of 1,415 BC and a mean IA date of 5,353 BC. Interestingly, the MRCAs are nearly always found in eastern Asia. This is due to the proximity of this region to both Eurasia and either the remote Pacific islands or the Americas, allowing the MRCA's descendants to reach a few major world regions in a relatively short time.

Arguably, this simulation is far too conservative, especially given its prediction that, even in densely populated Eurasia, only 55.3 people will leave each country per generation in AD 1500. If the migration rate among towns is increased to 20%, the local port users are reduced to 80%, and the migration rates between countries and continents are scaled up by factors of 5 and 10, respectively, the mean MRCA date is as recent as AD 55 and the mean IA date is 2,158 BC. The predictions of the simple ten-node graph model sketched earlier fall somewhere between these dates and those of the more conservative computational model.

The model also can be used to calculate the percentage of ancestry that current individuals receive from different parts of the world. In generations sufficiently far removed from the present, some ancestors appear much more often than do others on any current individual's family tree, and can therefore be expected to contribute proportionately more to his or her genetic inheritance[1,10,11]. For example, a present-day Norwegian generally owes the majority of his or her ancestry to people living in northern Europe at the IA point, and a very small portion to people living throughout the rest of the world. Furthermore, because DNA is inherited in relatively large segments from ancestors, an individual will receive little or no actual genetic inheritance from the vast majority of the ancestors living at the IA point[12].

Several factors could cause the time to the true MRCA or IA point to depart from the predictions of our model. If a group of humans were completely isolated, then no mixing could occur between that group and others, and the MRCA would have to have lived before the start of the isolation. A more recent MRCA would not arise until the groups were once again well integrated. In the case of Tasmania, which may have been completely isolated from mainland Australia between the flooding of the Bass Strait, 9,000–12,000 years ago, and the European colonization of the island, starting in 1803 (ref. 13), the IA date for all living humans must fall before the start of isolation. However, the MRCA date would be unaffected, because today there are no remaining native Tasmanians without some European or mainland Australian ancestry.

No large group is known to have maintained complete reproductive isolation for extended periods. The populations on either side of the Bering Strait appear to have exchanged mates throughout the period documented in the archaeological record[14]. Religious isolates such as the Samaritans occasionally have absorbed migrants from outside the group[15]. Even populations on isolated Pacific islands have experienced occasional infusions of newcomers[16]. Even if rates of migration between some adjoining populations are very low, the time to the MRCA tends not to change substantially. For example, with a migration rate across the Bering Strait of just one person in each direction every ten generations, rather than the ten per generation in the more conservative simulation described earlier, T_n only increases from 3,415 years to 3,668 years.

Conversely, other factors could reduce the time to the MRCA from that predicted by the model. Examples of such factors include the existence of more diverse intercontinental migration routes, the large-scale movement and mixing of populations documented in the historical record[17], marked individual differences in fertility[18], and the population increase of the past two millennia, which would result in more migrants.

Actual migration rates among populations are very poorly known and undoubtedly have varied considerably in different times and places. Studies of hunter-gatherer groups and subsistence agricultural communities have found that anywhere from 1% (ref. 19) to as much as 30% (ref. 20) of mates are from outside the group. The tendency of most human groups to marry out with surrounding groups, at least to a limited extent, links networks of ancestry within specific regions (see http://www.compapp.dcu.ie/~humphrys/FamTree/Royal/Famous.descents.html).

Given the remaining uncertainties about migration rates and real-world mating patterns, the date of the MRCA for everyone living today cannot be identified with great precision. Nevertheless, our results suggest that the most recent common ancestor for the world's current population lived in the relatively recent past—perhaps within the last few thousand years. And a few thousand years before that, although we have received genetic material in markedly different proportions from the people alive at the time, the ancestors of everyone on the Earth today were exactly the same.

Further work is needed to determine the effect of this common ancestry on patterns of genetic variation in structured populations[21–24]. But to the extent that ancestry is considered in genealogical rather than genetic terms, our findings suggest a remarkable proposition: no matter the languages we speak or the colour of our skin, we share ancestors who planted rice on the banks of the Yangtze, who first domesticated horses on the steppes of the Ukraine, who hunted giant sloths in the forests of North and South America, and who laboured to build the Great Pyramid of Khufu. □

Received 30 December 2003; accepted 14 July 2004; doi:10.1038/nature02842.

1. Wachter, K. W. in *Genealogical Demography* (eds Dyke, B. & Morrill, W. T.) 85–93 (Academic, New York, 1980).
2. Chang, J. T. Recent common ancestors of all present-day individuals. *Adv. Appl. Probab.* **31,** 1002–1026, 1027–1038 (1999).
3. Derrida, B., Manrubia, S. C. & Zanette, D. H. On the genealogy of a population of biparental individuals. *J. Theor. Biol.* **203,** 303–315 (2000).
4. Ingman, M., Kaessmann, H., Pääbo, S. & Gyllensten, U. Mitochondrial genome variation and the origin of modern humans. *Nature* **408,** 708–713 (2000).
5. Thomson, R., Pritchard, J. K., Shen, P., Oefner, P. J. & Feldman, M. W. Recent common ancestry of human Y chromosomes: Evidence from DNA sequence data. *Proc. Natl Acad. Sci. USA* **97,** 7360–7365 (2000).
6. Hudson, R. R. in *Oxford Surveys of Evolutionary Biology* (eds Harvey, P. H. & Partridge, L.) 1–44 (Oxford Univ. Press, New York, 1990).
7. Stannard, D. E. *American Holocaust: Columbus and the Conquest of the New World* (Oxford Univ. Press, New York, 1992).
8. US National Office of Vital Statistics, *Death Rates by Age, Race, and Sex, United States, 1900–1953, Vital Statistics—Special Reports* **Vol. 43** (US Government Printing Office, Washington DC, 1956).
9. Pletcher, S. D. Model fitting and hypothesis testing for age-specific mortality data. *J. Evol. Biol.* **12,** 430–439 (1999).
10. Ohno, S. The Malthusian parameter of ascents: What prevents the exponential increase of one's ancestors? *Proc. Natl Acad. Sci. USA* **93,** 15276–15278 (1996).
11. Derrida, B., Manrubia, S. C. & Zanette, D. H. Distribution of repetitions of ancestors in genealogical trees. *Physica A* **281,** 1–16 (2000).
12. Wiuf, C. & Hein, J. On the number of ancestors to a DNA sequence. *Genetics* **147,** 1459–1468 (1997).
13. Jones, R. Tasmanian archaeology: Establishing the sequences. *Ann. Rev. Anthropol.* **24,** 423–446 (1995).
14. Fitzhugh, W. W. & Chausonnet, V. (eds) *Crossroads of Continents: Cultures of Siberia and Alaska* (Smithsonian Institution Press, Washington DC, 1988).
15. Bonné-Tamir, B. *et al.* Maternal and paternal lineages of the Samaritan isolate: Mutation rates and time to most recent common male ancestor. *Ann. Hum. Genet.* **67,** 153–164 (2003).
16. Morton, N. E., Harris, D. E., Yee, S. & Lew, R. Pingelap and Mokil atolls: Migration. *Am. J. Hum. Genet.* **23,** 339–349 (1971).
17. Hoerder, D. *Cultures in Contact: World Migrations in the Second Millennium* (Duke Univ. Press, Durham, North Carolina, 2002).
18. Zerjal, T. *et al.* The genetic legacy of the Mongols. *Am. J. Hum. Genet.* **72,** 717–721 (2003).
19. Weiss, K. M. & Maruyama, T. Archeology, population genetics and studies of human racial ancestry. *Am. J. Phys. Anthropol.* **44,** 31–50 (1976).
20. Ward, R. H. & Neel, J. V. Gene frequencies and microdifferentiation among the Makiritare indians. IV. A comparison of a genetic network with ethnohistory and migration matrices; a new index of genetic isolation. *Am. J. Hum. Genet.* **22,** 538–561 (1970).

21. Jorde, L. B. in *Current Developments in Anthropological Genetics* (eds Mielke, J. H. & Crawford, M. H.) 135–208 (Plenum, New York, 1980).
22. Notohara, M. The coalescent and the genealogical process in geographically structured populations. *J. Math. Biol.* **29,** 59–75 (1990).
23. Wilkinson-Herbots, H. M. Genealogy and subpopulation differentiation under various models of population structure. *J. Math. Biol.* **37,** 535–585 (1998).
24. Hey, J. & Machado, C. A. The study of structured populations—new hope for a difficult and divided science. *Nature Rev. Genet.* **4,** 535–543 (2003).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements The research of D.L.T.R. was supported by the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.L.T.R. (dr@tedlab.mit.edu).

Phenotypic consequences of 1,000 generations of selection at elevated CO_2 in a green alga

Sinéad Collins & Graham Bell

Biology Department, McGill University, Montreal, Quebec H3A 1B1, Canada

Estimates of the effect of increasing atmospheric CO_2 concentrations on future global plant production rely on the physiological response of individual plants or plant communities when exposed to high CO_2 (refs 1–6). Plant populations may adapt to the changing atmosphere, however, such that the evolved plant communities of the next century are likely to be genetically different from contemporary communities[7–12]. The properties of these future communities are unknown, introducing a bias of unknown sign and magnitude into projections of global carbon pool dynamics. Here we report a long-term selection experiment to investigate the phenotypic consequences of selection for growth at elevated CO_2 concentrations. After about 1,000 generations, selection lines of the unicellular green alga *Chlamydomonas* failed to evolve specific adaptation to a CO_2 concentration of 1,050 parts per million. Some lines, however, evolved a syndrome involving high rates of photosynthesis and respiration, combined with higher chlorophyll content and reduced cell size. These lines also grew poorly at ambient concentrations of CO_2. We tentatively attribute this outcome to the accumulation of conditionally neutral mutations in genes affecting the carbon concentration mechanism.

Plant growth depends on CO_2 concentration[1,2], which is expected to rise from current levels of about 400 parts per million (p.p.m.) to between 700 and 1,000 p.p.m. during the next century[3]. In response, global plant productivity in forests[4], grasslands[5], agroecosystems[6] and other ecosystems is expected to increase. Projections of future net primary productivity are complicated by synchronous changes in temperature and other factors, but most models predict increases in the land–atmosphere and ocean–atmosphere fluxes from current values of >-2 petagrams (Pg) C per year to about -5 Pg C per year[3]. This process is likely to be complicated by shifts in the species composition of plant communities[7], and more fundamentally by evolutionary changes within plant populations. In the very long term, this may involve the extinction of some groups and the radiation of others[8], but within a few hundred generations most plant populations may adapt to the increased supply of inorganic carbon. Selection experiments with plants have demonstrated a variety of responses[9–12], but have been limited to fewer than ten generations. The long-term response to selection and the properties of populations adapted to elevated CO_2 remain unknown, and constitute an important limit on our ability to predict future plant productivity.

We used a microbial model system in which large population size and short generation time make it possible to evaluate evolutionary change caused by the spread of novel mutations over hundreds of generations. *Chlamydomonas reinhardtii* is a unicellular green alga that has been extensively used to study the physiology and genetics of photosynthesis[13]. It possesses a carbon-concentrating mechanism (CCM), which increases the concentration of CO_2 near the active site of ribulose 1,5-bisphosphate carboxylase–oxygenase (Rubisco), in common with most other eukaryotic microalgae that have been studied[14]. We set up ten isogenic selection lines from each of two ancestral genotypes, half being grown at ambient CO_2 (ambient lines) and half at a concentration that increased from ambient to 1,050 p.p.m. over about 600 generations and was then maintained at this level for a further 400 generations (high lines). At least 10^5 cells per line were transferred for 125 transfers in a buffered, nutrient-rich medium. The history of these lines thus emulates the conditions that photosynthetic organisms are likely to experience during the next century or so, with respect to CO_2 levels alone.

The physiological effect of elevated CO_2 concentration is expected to be an increase in photosynthesis, causing an increase in growth. Net photosynthesis in the ambient lines increased by about 30% when they were grown at high CO_2 (Fig. 1a). The ambient lines diverged through time so that by the end of the experiment they varied significantly in the rate of photosynthesis (one-way analysis of variance (ANOVA): $F_{9,18} = 9.0$, $P < 0.001$) when grown at ambient CO_2 concentrations. The high lines had normal rates of photosynthesis at ambient CO_2, which increased by more than 50% as an average over all lines at high CO_2. However, this effect was very inconsistent: one group of high lines had low rates whereas a second group had very high rates of photosynthesis at high CO_2 concentration (Fig. 1a). This distinction was not related to the identity of the ancestor, and represented significantly more divergence in photosynthetic rates than was seen in the ambient lines ($F_{1,16} = 10.5$, $P = 0.005$).

The growth rate of cultures grown at elevated CO_2 was correlated with their photosynthetic rate among the ambient lines, but not among the high lines (Fig. 1b). The physiological effect of CO_2 on photosynthesis was reflected by growth in pure culture, where the maximal rate of increase (Fig. 1c) and the limiting density (Fig. 1d) of both the ambient and the high lines are enhanced substantially by high CO_2. However, there was no indication of a parallel evolutionary response: by the end of the selection experiment, the high lines had not become specifically adapted to growth at high CO_2; their growth at high CO_2 being no greater than, and perhaps even less than, the growth of the ambient lines. There was nevertheless an indirect response: the growth of some high lines was markedly impaired at ambient CO_2 concentrations where two of the lines could scarcely be propagated. This result was supported by the outcome of competition assays in which the selection lines were mixed with standard genetically marked strains and the change in frequency during growth in culture recorded (Table 1). The high lines had considerably lower competitive ability at ambient CO_2, where three of them (including the two with strongly reduced growth in pure culture) were such weak competitors that they were consistently eliminated by the tester strains within 10–15 generations. They were, however, no more successful than the ambient lines at high CO_2. In short, 1,000 generations of selection at high CO_2 concentrations had caused no increase in growth at high CO_2, whereas growth at ambient CO_2 was often considerably reduced.

Photosynthesis is linearly related to respiration in the dark among lines at ambient CO_2; this relationship is the same for ambient and high lines (Fig. 2a). It has been shown in *Chlamydomonas* that post-illumination rates of O_2 consumption provide a