

- overhang was filled with Klenow.
38. S. M. Beverley and C. E. Clayton, in *Protocols in Molecular Parasitology*, J. E. Hyde, Ed. (Humana, Totowa, NJ, 1993), pp. 333–348.
 39. A targeting segment of pHD404 (black box in Fig. 3) corresponding to the region upstream of the PARP promoter was generated by polymerase chain reaction (PCR) with cloned PARP A locus DNA as the template and the primers TACGGTACCGCGGCCG-CAACAAAGGCTGACTGCACCT and TACGAGCT-CTTTCAGACTTCTTCGATGCC (introduced Kpn I, Not I, and Sac I restriction sites are underlined).
 40. To construct a PARP promoter with internal Kpn I and Not I sites, we cloned the PARP A upstream targeting segment into pHD63 cut with Kpn I and Sac I (29). This was transferred as a Sac I-Xho I fragment to pHD330 (12), replacing the tubulin homology and yielding pHD383. The hygromycin resistance gene of pHD383 was then replaced by the phleomycin resistance gene of pHD63, producing pHD403. Finally the Spe I-Bam HI fragment of pHD403 spanning the downstream region of the wild-type PARP promoter and CAT gene was replaced with the Spe I-Bam HI fragment of pHDTop4, supplying the operator-bearing PARP promoter, splice acceptor, and luciferase gene and yielding pHD404.
 41. Clone 3 trypanosomes expressing the TetR (see Fig. 1) were transfected with 10 μ g of Not I-linearized pHD404 or pHD430 (per 10^7 cells per cuvette). After overnight culture in hygromycin (50 μ g/ml) and Tc (10 μ g/ml), phleomycin was added to a final concentration of 5 μ g/ml, and trypanosomes were diluted serially in microtitre plates.
 42. This fragment was generated by PCR by using a cloned fragment from the rRNA locus intergenic region as template (28) and primers GATCTC-GAGATGGGTACCGGTGTGTGCCAAAGACATT

and GCAGAGCTCAGGCTTTCGGACATGAATTG (introduced Xho I, Kpn I, and Sac I sites are underlined); a Not I linker was inserted into the unique Cla I site.

43. We thank H. Bujard, F.-B. Wang, A. Bonin, and M. Gossen (Zentrum für Molekulare Biologie der Universität Heidelberg) for the plasmid pUHD14-1 and for helpful discussions; S. Beverley (Harvard) for discussions and communicating unpublished results; G. Rudenko (Netherlands Cancer Institute) for calling our attention to the rDNA nontranscribed spacer as a potential integration site and for providing plasmids containing cloned fragments from the rRNA locus; G. Hobbs for grammatical insight; and J. Zutt for the serene environment. This work was partially supported by the Deutsche Forschungsgemeinschaft (SFB 229 Teil D5).

25 October 1994; accepted 10 February 1995

Absence of Polymorphism at the ZFY Locus on the Human Y Chromosome

Robert L. Dorit,* Hiroshi Akashi, Walter Gilbert

DNA polymorphism in the Y chromosome, examined at a 729-base pair intron located immediately upstream of the ZFY zinc-finger exon, revealed no sequence variation in a worldwide sample of 38 human males. This finding cannot be explained by global constraint on the intron sequence, because interspecific comparisons with other nonhuman primates revealed phylogenetically informative sequence changes. The invariance likely results from either a recent selective sweep, a recent origin for modern *Homo sapiens*, recurrent male population bottlenecks, or historically small effective male population sizes. A coalescence model predicts an expected time to a most recent common ancestral male lineage of 270,000 years (95 percent confidence limits: 0 to 800,000 years).

The human Y chromosome, nonrecombining along most of its length and paternally inherited, should be extremely useful for the reconstruction of genetic and evolutionary history. However, relatively little is known about the patterns of polymorphism between human Y chromosomes. Polymorphism has been reported for certain regions of the chromosome; with few exceptions (1, 2) these reports involve either the use of anonymous probes of uncharacterized sequence (3) or represent variable numbers or locations of Alu or other repetitive elements (4).

Here we report a study of sequence variation at a well-characterized human Y-linked locus: a 729-base pair (bp) intron located between the third exon and the zinc-finger-encoding fourth exon of the ZFY locus (5). This gene, located in interval Yp-1A2, is actively transcribed in males and appears to be involved in sperm or testes maturation (6). We carried out a detailed survey of this region in a worldwide sample of humans in order to provide a preliminary

picture of sequence polymorphism on the Y chromosome. We surveyed 38 individuals, chosen to represent a cross section of geographic origins, and sequenced the entire intron in all of them (7). In addition, we sequenced part or all of the 3'-most zinc-finger exon in 12 of those individuals, as well as the homologous intron in three other nonhuman primates—chimpanzee, gorilla, and orangutan (8).

Surprisingly, we detected no intraspecific polymorphism whatsoever, in either the intron or the exon, in our human sample. Such an absence of variation across the 729-bp intron in a sample of this size (a total of ~28,000 bp sampled) is unexpected, because intron sequences appear to be subjected to few sequence-specific constraints. Selection at this intron cannot account for the absence of variation, as interspecific comparisons of the sequences of this intron in other primates show that variable sites are distributed throughout the intron and include at least 21 unambiguous transitions, 14 unambiguous transversions, and 4 insertions or deletions (8). Furthermore, these data suggest that the absence of recombination in this region of the Y chromosome does not detectably slow rates of interspecific divergence; indeed, Y-linked sequences have been shown to exhibit accelerated rates of evolution (9).

Table 1 summarizes the interspecific differences at this region, which were then used to construct a parsimony tree of these sequences (Fig. 1). These Y chromosome sequences yielded a completely unambiguous shortest-length tree (Fig. 1) uniting the human and chimpanzee sequences on the basis of two synapomorphic changes (10).

The interspecific data predict a level of Y chromosome polymorphism in *Homo sapiens*, assuming clock-like behavior of these sequences. The divergences listed in Table 1 correspond to an approximate mean mutation fixation rate of 0.135% per million years (My) of elapsed time along a single branch for this intron (11). Given the total human intron sequence examined in this study (28,000 bp), we would expect 19.5 segregating sites if human lineages were to trace back 500,000 years on average. If the origin of modern *Homo sapiens* is more recent (on the order of 150,000 years), the expectation for segregating sites declines to 5.5.

There are three general classes of explanation for the lack of variation of a chromosomal region: purifying selection, chance absence of segregating sites, or recent common ancestry. The interspecific comparisons mentioned previously rule out a selectively mediated global conservation of the intron sequence. Could the sampling variation in the distribution of polymorphic sites among individuals produce the observed monomor-

Table 1. Absolute (below diagonal) and mean (above diagonal) interspecific distances for the final ZFY intron, averaged over all possible states. Distances calculated according to the Kimura two-parameter model (18) are listed in parentheses. Mean sequence length compared = 729 bp.

	Human	Chimp	Gorilla	Orang
Human		0.007 (0.007)	0.014 (0.013)	0.041 (0.042)
Chimp	5		0.015 (0.015)	0.043 (0.043)
Gorilla	10	11		0.044 (0.046)
Orang	30	31	32	

R. L. Dorit, Department of Biology, Yale University, New Haven, CT 06511, USA.
H. Akashi, Ecology and Evolution Department, University of Chicago, Chicago, IL 60637, USA.
W. Gilbert, Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA.

*To whom correspondence should be addressed.

phism, given an overall expectation for human nucleotide polymorphism? Stated otherwise, what lowest actual human sequence diversity (Θ) rules out sampling alone as the cause of the observed monomorphism? If we use a very conservative test (12), coalescence theory (assuming random mating in a population of constant size) argues that any value of $\Theta \geq 0.0011$ predicts polymorphism in our sample (with $P \geq 0.95$). Although a clear picture of nucleotide diversity in *Homo sapiens* has yet to emerge, Table 2 summarizes the data for several autosomal regions examined to date. The critical value for this study thus falls below most, but not all, available estimates, thus suggesting that the lack of polymorphism at ZFY is not due to chance (13).

A very recent common ancestry for this region of the Y chromosome may reflect the influence of demographic history and of certain population genetic parameters, including small effective male population size. The absence of recombination along most of the Y chromosome exaggerates the impact of such parameters, since the entire chromosome, excluding the pseudoautosomal region (PAR), is in permanent linkage disequilibrium and behaves as a single linkage group. A rapid, selectively driven fixation of an allele at any locus within the linkage group would result in the loss of polymorphism at all linked loci and would obliterate any evi-

dence of older ancestry for all Y chromosome lineages. The resulting fixed chromosome thus becomes the last common ancestor for all subsequent Y chromosome lineages. Such a fixation event must necessarily have been recent, because no polymorphism has subsequently accumulated—at least in the region sampled—as a result of mutational events following the selective sweep (14). We can estimate the age of that common ancestor by investigating the elapsed time consistent with the absence of polymorphism we observe. A coalescent model, with its assumptions of random mating, equilibrium population size, and exponentially distributed bifurcation times, provides an expected date for the last common male ancestor of 270,000 years (with 95% confidence limits of 0 to 800,000 years). Increasing the population size or nonrandom mating would lower this estimate. A lowest limit for the age of the last common ancestor of all Y lineages is derived by assuming the rapid branching and subsequent independence of all Y lineages since the last common male ancestor (known as a “star” phylogeny); such a pattern provides an estimate of 27,000 years, with 95% limits of 0 to 80,000 years. A mixed model, involving local (regional) coalescence, would produce intermediate times (15).

Is this pattern of monomorphism a locus-, region-, or chromosome-wide phenomenon? Low-level variation has been reported in the form of restriction fragment length polymorphisms (RFLPs) for the Y chromosome detected by means of a variety of anonymous probes (3), most of them directed to the nonrecombining region Yq11. These RFLPs, for the most part, reflect CpG loss or possible duplication. Geographically structured polymorphism in the number and location of Alu insertions has also been suggested (2, 4). A very low level of Y chromosome polymorphism thus appears to have been maintained or restored.

We note that the results presented here are not compatible with most multiregional models for the origin of modern humans (16). While the geographic location of the

ancestral Y lineage cannot be determined from these data, the age of that lineage does lend support to the hypothesis of a recent, single origin of modern *Homo sapiens*, and one of our estimates (270,000 years) is congruent with the estimated age of the ancestral mitochondrial DNA lineage (17).

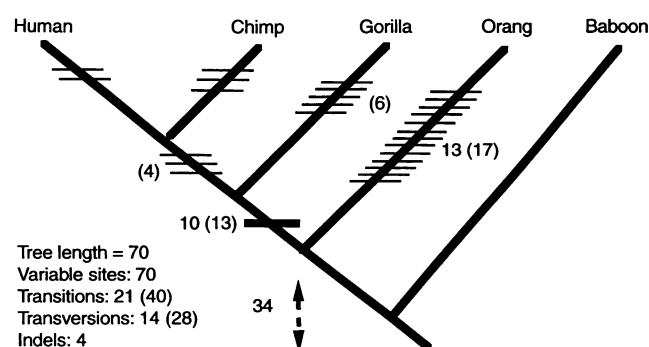
REFERENCES AND NOTES

1. N. Ellis *et al.*, *Nature* **344**, 663 (1990).
2. L. Roewer, J. Arnemann, N. K. Spurr, K.-H. Grzechick, J. T. Epplen, *Hum. Genet.* **89**, 389 (1992).
3. G. Lucotte and N. Y. Ngo, *Nucleic Acids Res.* **13**, 8285 (1985); K. Y. Ngo, G. Vergnaud, C. Johnsson, G. Lucotte, J. Weissenbach, *Am. J. Hum. Genet.* **38**, 407 (1986); G. Lucotte, in *The Human Revolution: Behavioral and Biological Perspectives on the Origins of Modern Humans*, P. Mellars and C. Stringer, Eds. (Edinburgh Univ. Press, Edinburgh, 1988), pp. 39–46; A. Torroni *et al.*, *Ann. Hum. Genet.* **54**, 287 (1990); R. Oakey and C. Taylor-Smith, *Genomics* **7**, 325 (1990); P. Malaspina *et al.*, *Ann. Hum. Genet.* **54**, 297 (1990); A. B. Spurdle and T. Jenkins, *Am. J. Hum. Genet.* **50**, 107 (1991).
4. M. F. Hammer, *Mol. Biol. Evol.* **11**, 749 (1994); A. B. Spurdle, M. F. Hammer, T. Jenkins, *Am. J. Hum. Genet.* **54**, 319 (1994).
5. D. C. Page *et al.*, *Cell* **51**, 1091 (1987); M. S. Palmer *et al.*, *Nature* **342**, 937 (1989).
6. A. Scheiner-Gadicke, P. Beer-Romero, L. G. Brown, R. Nussbaum, D. C. Page, *Cell* **57**, 1247 (1989); Y.-F. C. Lau and K. Chan, *Am. J. Hum. Genet.* **45**, 942 (1989); C. Nagamine, K. Chan, E. Hakel, Y.-F. C. Lau, *Genes Dev.* **4**, 63 (1990); H. Su and Y.-F. C. Lau, *Mol. Reprod. Dev.* **33**, 252 (1992).
7. Human DNA samples were obtained from male volunteers who donated hair follicle samples or from cell lines provided by L. L. Cavalli-Storza and K. K. Kidd. Geographic origins were determined by interview. Whenever possible, geographic origins of parents and grandparents were also ascertained. The samples are grouped by continent of origin, and the number of individuals is given in parentheses. Africa: Nigeria* (1), Ivory Coast (1), Tanzania (1), Southern Africa (2), Algeria (1), Central African Republic* (2), African American (2); Americas: Mexico (2), Guatemala (1), Peru* (1), Argentina (1), Native American (2); Asia: China* (2), Korea (1), Japan* (2), Taiwan (2), Indonesia (1), India (1); Europe/Middle East: Ireland* (1), Belgium (1), Italy* (1), Spain (1), Russia* (2), Poland* (1), Saudi Arabia* (1), Turkey (1); South Pacific: Melanesia (1), New Guinea* (1), Australia* (1). (*) Indicates samples where the 3'-most zinc-finger exon was also sequenced.
8. Polymerase chain reaction (PCR) primers were designed on the basis of the sequence of the exon 3' and 5' to this intron (upstream: 5'-CTTGGCAGAGTGGCTAAACAG-3'; downstream: 5'-CCTCGACTTAAACTTCTTCCC-3'; the zinc-finger exon, when needed, was obtained by replacing the above downstream primer with the following: 5'-ACTTCTTATGATGTCGCAT-3'). Whole genomic DNA was prepared either from crude preparations of cheek or hair root cells obtained from volunteers, or from genomic DNA preparations of sperm cells and was then amplified by PCR (93°C, 1 min; 55°C, 1 min; 72°C, 1 min; 35 cycles). The analogous intron from the ZFX locus, occasionally amplified in our protocols, was significantly larger (1.2 versus 0.8 kb), because of the presence of an inserted Alu element in the ZFX intron. The PCR amplifications yielded products that were clearly visualized on an ethidium-stained agarose gel, and the Y-derived ZFY intron could always be identified and separated. A negative control was included in all amplifications to ensure the absence of contamination; in addition, a nonhuman primate sample was included as a positive control and to ensure that no template cross-contamination took place. All samples were amplified at least three times in separate experiments. The amplification products were then sequenced directly by Maxam-Gilbert chemical sequencing, genomic sequencing, or dideoxy-sequencing of double- or single-stranded PCR product. The sequences are deposited in GenBank (accession

Table 2. Sample autosomal nucleotide diversity estimates for human populations.

Locus or region	Nucleotide diversity (Π)	Ref.
Alpha-1-antitrypsin	0.0021 (Baltimore)	19
Alpha-1-antitrypsin	0.0010 (Houston)	19
Beta-globin	0.004	20
17 3'-regions (Ch 21)	0.00098	21
Growth hormone gene	0.002	22
Apolipoprotein	0.0021	23
Insulin	0.0017	24
Composite (Θ)	0.0011	25
Cosmid 4p16.3	0.0023	26

Fig. 1. Most parsimonious phylogenetic network uniting the ZFY intron sequences described in this study. Horizontal bars indicate unambiguous sequence changes along a particular branch, numbers in parentheses indicate maximum numbers of changes along the branch. Changes below the node uniting the human, chimp, gorilla, and orang sequences cannot be assigned unambiguously. Network was constructed with the exhaustive search algorithm in PAUP and rooted with the baboon sequence as an outgroup (sequence provided by W.-H. Li).



- numbers are U24117 (chimp), U24118 (human), U24119 (gorilla), and U24120 (orang); alignments are available upon request.
9. See also L. C. Shimmin, H.-J. Chang, W.-H. Li, *Nature* **362**, 745 (1993). Our sequences for the human ZFY intron differ from that reported by Shimmin *et al.* at position 478, which represents the final T in a 12 T homopolymeric run. This may represent either a legitimate polymorphism in this region or more likely a sequencing discrepancy.
 10. The topology of the most parsimonious tree was derived by exhaustive search with the PAUP package. The tree was rooted with the baboon sequence as an outgroup. The tree shown is 70 steps long [CI (consistency index) = 0.986, RI (retention index) = 0.923]; the next shortest tree, 72 steps long, results in a chimp-gorilla-human trichotomy.
 11. This value was determined by using 14 My as the branching date for the orangutan lineage and 5 My as the age of the human/chimp-gorilla split, as proposed by M. Hasegawa, H. Kishino, and T. A. Yano [*J. Hum. Evol.* **18**, 461 (1989)]. The rate was derived by averaging over all pairwise comparisons.
 12. M. Kreitman and R. R. Hudson, *Genetics* **127**, 565 (1991); R. R. Hudson, in *Mechanisms of Molecular Evolution*, N. Takahata and A. G. Clark, Eds. (Sinauer, Sunderland, MA, 1993, pp. 23–36).
 13. The expected nucleotide diversity of Y chromosome loci may be reduced by the smaller effective population size of the Y chromosome (1/4 N_e relative to autosomal loci)—a given value of X thus leads to a fourfold reduction in predicted polymorphism relative to an autosomal locus. This effect, however, is offset by the apparently increased mutation rate for Y-linked sequences (9).
 14. A. J. Berry, J. W. Ajioka, M. Kreitman, *Genetics* **129**, 1111 (1991).
 15. We used a coalescence theory approach to estimate the probability of finding zero mutations in a current sample of 38 Y chromosomes as a function of the expected mutation rate, μ , and the deepest branching time, T . This probability is

$$p(0|T) = \prod_{i=1}^{37} (i(i + \mu T))$$

where μ was estimated from the chimp, gorilla, and orangutan data as 0.135% per site per million years; given 729 sites, $\mu = 0.98$ per million years. We convert this to a Bayesian expectation for the time, given zero mutations, as

$$p(T|0) = p(0|T)p(T)/p(0) = p(0|T) \int_0^T p(0|t) dt$$

and computed both an expected value for the time, T , as

$$\langle T \rangle = \int_0^T p(0|t) dt \int_0^T p(0|t) dt$$

and a 95% maximum estimate for the time as T_m from

$$\int_0^{T_m} p(0|t) dt = 0.95 \int_0^{\infty} p(0|t) dt$$

This yielded the values $\langle T \rangle = 270,000$ years and $T_m \sim 800,000$ years. For a population of N Y chromosomes, $T = 2 \times$ generation time $\times N$. Thus for a 20-year generation time, the coalescence picture is one of an effective population of 7500 males. This is an exceedingly small population size for this entire 300,000 year period; it is far more likely that the coalescence model, which assumes worldwide uniform mixing and a constant effective population size, is not strictly applicable, and that humans recently fanned out around the world into groups which remained partially isolated after that time. That picture is more similar to the "star" phylogeny, each of the 38 chromosomal lineages remaining distinct after an original migration; such a star phylogeny corresponds to a $\langle T \rangle$ of 27,000 years for that original "migration."

16. C. B. Stringer and P. Andrews, *Science* **239**, 1263 (1988); M. H. Wolpoff, in *The Human Revolution, Behavioral and Biological Perspectives on the Origins of Modern Humans*, P. Mellars and C. Stringer, Eds. (Princeton Univ. Press, Princeton, NJ, 1989), pp. 62–108; D. W. Frayer, M. H. Wolpoff, A. G. Thorne, F. H. Smith, G. G. Pope, *Am. Anthropol.* **95**, 14 (1993); F. J. Ayala, A. Escalante, C. O'Huigin, J. Klein, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 6786 (1994).
17. L. Vigilant, M. Stoneking, H. Harpending, K. Hawkes, A. C. Wilson, *Science* **253**, 1503 (1991); M. Hasegawa and S. Horai, *J. Mol. Evol.* **32**, 37 (1991).
18. M. Kimura, *J. Mol. Evol.* **16**, 111 (1980).
19. K. J. Matteson *et al.*, *Hum. Genet.* **69**, 263 (1985).
20. A. Chakravarti, K. H. Buetow, S. E. Antonarakis, C. D. Boehm, H. H. Kazazian, *Am. J. Hum. Genet.* **33**, 134a (1981).
21. D. Avramopoulos, A. Chakravarti, S. Antonarakis, *Genomics* **15**, 98 (1993).
22. A. Chakravarti, J. A. Phillips III, K. H. Mellits, K. H. Buetow, P. H. Seeburg, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 6085 (1984).
23. S. Antonarakis *et al.*, *Hum. Genet.* **80**, 265 (1988).
24. A. Chakravarti, S. C. Elbein, M. A. Permutt, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 1045 (1986).
25. W.-H. Li and L. A. Sadler, *Genetics* **129**, 513 (1991).
26. R. McCombie *et al.*, *Nat. Genet.* **1**, 348 (1992).
27. We thank D. Page for encouragement and C. Lowen for technical assistance. Supported by NIH grant GM 37997 to W.G. and by the Yale Science Development Fund.

26 October 1994; accepted 16 March 1995

T Helper Cell Subsets in Insulin-Dependent Diabetes

Jonathan D. Katz,* Christophe Benoist, Diane Mathis†

It has been proposed that the development of insulin-dependent diabetes is controlled by the T helper 1 (T_H1) versus T_H2 phenotype of autoreactive T_H cells: T_H1 cells would promote diabetes, whereas T_H2 cells would actually protect from disease. This proposition was tested by establishing cultures of T_H1 and T_H2 cells that express an identical diabetogenic T cell receptor and comparing their ability to initiate disease in neonatal nonobese diabetic mice. T_H1 -like cells actively promoted diabetes; T_H2 -like cells invaded the islets but did not provoke disease—neither did they provide substantial protection.

Insulin-dependent diabetes mellitus is an autoimmune disease characterized by infiltration of leukocytes into the islets of Langerhans of the pancreas and breakdown of glucose homeostasis as a result of destruction of insulin-producing beta cells (1). The leukocyte infiltrate, termed insulinitis, is a heterogeneous population, composed of $CD4^+$ and $CD8^+$ T lymphocytes, B lymphocytes, macrophages, and dendritic cells (2). T lymphocytes play a primary role (3), but the relative contribution of $CD4^+$ and $CD8^+$ cells is uncertain, in particular the importance of each subset in provoking or perpetuating disease and whether each has a unique function (4). Nonetheless, there have been several reports of $CD4^+$ T cells, by themselves, instigating diabetes (5–7).

$CD4^+$ T lymphocytes fall into two major classes: T helper 1 (T_H1) cells, which secrete interferon γ (IFN- γ) and are primarily associated with cellular immunity; and T_H2 cells, which produce interleukin-4 (IL-4) and IL-10 and are mainly involved in humoral immunity (8). Several studies have correlated diabetes with T_H phenotype, leading to the idea that T_H1 cells promote disease whereas T_H2 cells protect from it, dampening the activity of T_H1 effectors (9). The most sub-

stantive arguments in support of this notion derive from studies in mice. Artificial introduction of lymphokines or antibodies to lymphokines that favor T_H1 or disfavor T_H2 cell development generally promotes diabetes, whereas the converse inhibits disease (10). Along similar lines, C57Bl mice are more likely to develop a T_H1 and less likely to develop a T_H2 response than are BALB/c mice (11), and transgenic animals that express a T cell receptor (TCR) reactive to an islet cell neoantigen develop diabetes on the former but not the latter genetic background (12). Finally, there appears to be a correlation between the amount and type of lymphokines secreted by invading lymphocytes and how aggressively the lymphocytes attack an islet graft, high IFN- γ and low IL-4 levels being associated with destruction (13).

However, the above set of arguments appears to be inconclusive for three major reasons. First, conflicting data do exist. For example, pancreatic expression of IL-10, a T_H2 lymphokine, actually promotes disease in nonobese diabetic (NOD) mice (14). Second, many of the points cited above rely on effects provoked by manipulating the levels of particular lymphokines; however, lymphokines are highly pleiotropic and control many phenomena besides T_H1 and T_H2 phenotype: the differentiation of $CD8^+$ T cells, B cell differentiation and effector diversification, natural killer cell activity, antigen-presenting cell function, and lymphocyte circulation. In addition, beta islet cells appear

Institut de Génétique et de Biologie Moléculaire et Cellulaire, INSERM-CNRS-Université Louis Pasteur, BP 163, 67404 Illkirch Cédex, CU de Strasbourg, France.

*Present address: Department of Pathology, Washington University, St. Louis, MO 63110, USA.

†To whom correspondence should be addressed.