

LETTER TO JMG

Estimating the age of rare disease mutations: the example of Triple-A syndrome

E Genin, A Tullio-Pelet, F Begeot, S Lyonnet, L Abel

J Med Genet 2004;41:445–449. doi: 10.1136/jmg.2003.017962

Triple-A syndrome (MIM 231550) is an autosomal recessive disorder characterised by adrenocorticotrophin hormone resistant adrenal insufficiency, achalasia of the oesophageal cardia, and alacrima.¹ The gene, previously localised to chromosome 12q13,^{2,3} was recently identified and denoted as *AAAS*.⁴ Among the five homozygous truncating mutations that were characterised, a single splice donor splice mutation (IVS14+1G→A) was found in several unrelated affected individuals of north African origin, strongly suggesting a founder effect.⁴ In this work, we were interested in estimating the time at which the mutation occurred, since this is expected to provide interesting and helpful information on its natural history. Different methods have been proposed to estimate the age of mutations,^{5,6} which are based either on allele frequencies⁷ or on intra-allelic variability and the pattern of linkage disequilibrium at closely linked marker loci^{8,9} with extensions to the analysis of multilocus data.^{9–12} However, these latter approaches are dedicated more to the fine mapping of the mutation, and may not be appropriate for estimating the age of rare mutant alleles that are only found in very few affected individuals. Therefore, we developed a new simple method based on likelihood that uses multilocus marker data to estimate the age of the most recent common ancestor of the mutation from a small number of patients. The method was tested through simulations and then applied to nine consanguineous Triple-A patients, who were homozygous for the IVS14+1G→A mutation.

MATERIALS AND METHODS

Method of dating

We will first describe the principle of the method for two affected individuals who carry the studied mutation at the disease locus D , and then extend the method to a sample of N affected individuals. The basic assumption is that the two affected individuals descend from a common ancestor who introduced the mutation n_{gen} generations ago. The problem is to estimate n_{gen} from the size of the haplotype shared by the two individuals on each side of D . For simplicity, we will describe the method for only one side, the right side, as the treatment for the other side is completely analogous and independent. Let $(M_1, \dots, M_x, \dots, M_k)$, be a set of k ordered markers that have been typed on the right side of D , which are located at recombination fractions $(\theta_1, \dots, \theta_x, \dots, \theta_k)$ from D . M_k is the first marker on the right side for which the two individuals have different alleles. The problem can be understood as a survival analysis problem in which the starting point is the disease locus, the discrete time scale is the genetic distance, and the event of interest is the occurrence of a recombination. Therefore, we define:

- $S(x)$, the probability that no recombination took place during n generations between D and marker M_x :
 $S(x) = (1 - \theta_x)^n$

Key points

- Among the causing mutations that were recently characterised in Triple-A syndrome (MIM 231550), the IVS14+1G→A mutation was found in several unrelated affected individuals of north African origin, strongly suggesting a founder effect. Although different methods have been proposed to estimate the age of relatively common mutations, none of these methods has been evaluated in the context of rare monogenic diseases such as Triple-A syndrome.
- We first developed a simple likelihood based method to estimate the age of the most recent common ancestor of the patients, based on the information provided by haplotypes shared by affected individuals. Through simulation studies, we show that the method provided satisfactory results even with small samples (≈ 5 individuals), and investigate the influence on age estimates of misspecification of both allele frequencies and mutation rates at the marker loci.
- We applied the method to the analysis of nine patients with Triple-A, who were carriers of the IVS14+1G→A mutation, and found an age estimate of ≈ 1000 – 1175 years. Interestingly, this period corresponds to a time of important migrations into north Africa from the Arabian peninsula.

- $f(x)$, the probability that a recombination took place exactly in the x th interval between marker M_{x-1} and M_x :
 $f(x) = S(x-1) - S(x)$

For two individuals who shared all their alleles for markers (M_1, \dots, M_{k-1}) and have a different allele at marker M_k , two possibilities should be considered:

- for both individuals, a recombination did occur in the k th interval between markers M_{k-1} and M_k , and the likelihood is:

$$L_a(n) = f(k)^2 \quad (1)$$

- only one of the two individuals has recombined in the k th interval, and the other has received the ancestral allele at M_k (and thus not recombined by the k th interval); in this case the likelihood is:

$$L_b(n) = 2S(k)f(k) \quad (2)$$

Therefore, the likelihood for this side is the sum of $L_a(n)$ and $L_b(n)$, and the final likelihood for two individuals is the product of the two side specific likelihoods.

We consider now a sample of N affected individuals who carry the studied mutation, and assume that those N affected individuals descend independently from the same ancestor (star genealogy⁶). The issue of different lineages occurring at different times will be discussed later. For each side, two groups of individuals should be determined among the N . The first group, denoted as G_1 , corresponds to the y individuals ($2 \leq y \leq N$) who shared the longest haplotype—that is, those y individuals share all their alleles for markers (M_1, \dots, M_{k-1}), and have a different allele at marker M_k . The second group, denoted as G_2 , corresponds to the remaining $N-y$ individuals for which the first recombination occurred in an interval x_i , $i=1$ to $(N-y)$, closer to the mutation ($x_i < k$). For individuals of G_1 , the likelihood is written using the same principle as that for two individuals. Expressions 1 and 2 become

$$L_a(n) = f(k)^y \quad (3)$$

and

$$L_b(n) = yS(k)f(k)^{y-1} \quad (4)$$

The likelihood contribution of an individual i belonging to G_2 who has recombined within interval x_i is simply:

$$L_i(n) = f(x_i) \quad (5)$$

The likelihood of n given the data on the studied side is thus:

$$L_{side}(n) = [L_a(n) + L_b(n)] \times \prod_{i \in G_2} L_i(n) \quad (6)$$

and the final likelihood, $L(n)$, is the product of the two side specific likelihoods.

In all the previous derivations, we have not accounted for the possibility that a descendant haplotype may vary from its ancestor because of mutations rather than recombinations. This could lead to an underestimation of the age of the disease causing mutation to an extent that depends on the mutation rates at the markers. Moreover, equation 6 only holds if shared alleles indicate identity by descent so that the positions of crossovers can be inferred from the observed data without ambiguity. When the polymorphism of markers is low, it becomes necessary to take allele frequencies into account. We present in a supplementary material available at http://u535.vjf.inserm.fr/Pagesperso/GENIN/supplement_jmg.htm, extensions of the method with the corresponding formulas that take into account both mutations at the marker locus and marker allele frequencies. The supplementary material also provides the details of the computation for the 95% confidence interval of the estimated generation number. This computation is based on Bayesian principles as proposed in Piccolo et al.¹³

Simulated data

Simulations were conducted to investigate the proposed method. To generate individual haplotype data, we considered a mutation at a disease locus D and a set of equally distant markers on each side of D . For sample sizes larger than 10 individuals, two different recombination fractions between markers were considered: $\theta = 0.01$ and $\theta = 0.002$. For smaller sample sizes, to ensure that individuals shared at least one marker on each side of the mutation, distances between markers were reduced and the recombination

fractions were 0.0005. Markers were assumed to have different numbers of alleles with equal frequencies that were varied from 0.1 to 0.3. After simulating an ancestral haplotype and giving the number of generations, Monte Carlo methods were used to generate the positions of crossovers and marker mutations. The marker mutation rate, μ , was increased to 10^{-3} , since mutation rates of this magnitude have been observed for some short tandem repeat loci in human beings.¹⁴ This process was repeated until the desired number of individuals was reached. The simulated data were then analysed with the method described above. All computations were performed using a C program we developed, which is available upon request (email E Genin at genin@vjf.inserm.fr).

Different simulation conditions were considered according to the number of individuals per sample (from 2 to 30), number of generations (denoted as n_{gen} and ranging from 10 to 100), marker specificities (recombination fraction θ , allele frequency p , mutation rates). For each simulation condition, 5000 replicates were generated. The results present the mean number of generations (with empirical 95% confidence interval) estimated over the 5000 replicates. Because of the nature of the problem, the distribution of these estimates is highly skewed to the right, especially for low numbers of individuals and low numbers of generations. For this reason, we also report the median of the estimates, and the coverage—that is, the proportion of the replicates, p_{out-ci} , where the true value of n_{gen} falls outside the 95% confidence interval of the estimate.

Triple-A data

In the present study, we investigated nine consanguineous patients with clinical features defining Triple-A, who were homozygous for the IVS14+1G→A mutation. All patients originated from north Africa (mainly from Algeria and Tunisia) and had no known familial relationships. To estimate the age of the founder Triple-A mutation, we selected the polymorphic markers encompassing the AAAS locus according to the genetic and physical maps available at <http://genome.ucsc.edu> and www.ncbi.nlm.nih.gov. As the genetic distances available for closely linked markers are generally not very accurate, rates of recombination between markers were computed using both the overall genetic length of the haplotype and the physical distances between markers as proposed in Picard et al.¹⁵ The closest markers at which affected individuals do not share any alleles were determined on both sides of the AAAS locus. Genotyping was performed on DNA extracted from blood according to standard procedures as described in Hadj-Rabia et al.³ The frequency of shared alleles was estimated from a sample of 30 unrelated subjects of north African origin.

RESULTS

Simulation study

In the top of table 1, results are presented for a very small number of individuals (2 and 5) in a situation of complete identity by descent information and for a recombination fraction of 0.0005. Although an overestimation of the number of generations (n_{gen}) is observed for two individuals, this result is due to a small number of replicates that led to very high estimates of n_{gen} . In this situation, the median value provides a more reliable measure of the estimates, and is much closer to the true value of n_{gen} than the mean. When five individuals are used, the estimates become quite satisfactory for the median value as well as for p_{out-ci} . As the sample size increases (bottom of table 1), estimates become very good and confidence intervals are reduced. Similar results were obtained for a recombination of 0.002 between markers (data not shown), indicating that the

recombination fraction has no substantial influence on the estimates in the situation of complete information on identity by descent.

Table 2 presents the impact of misspecifying mutation rates in samples of 10 individuals for two recombination fractions ($\theta = 0.01$ and $\theta = 0.002$), and assuming complete information on identity by descent. Three different mutation rates (0 , 10^{-6} , and 10^{-3}) at the markers were used both to simulate (μ_{sim}) and analyse (μ_{ana}) the data. Table 2 presents the results when all allele changes are equally likely by mutation but the very same pattern of results (data not shown) is obtained when a stepwise mutation model is used (that is, a mutation can only modify the number of repeats by one). When the mutation rate is correctly specified ($\mu_{sim} = \mu_{ana}$, values in bold), the estimates of the numbers of generations are quite reliable and the coverage p_{out-ci} is very close to 5%. When mutations are introduced in the simulations but ignored in the analysis ($\mu_{ana} = 0$), the age of the disease mutation is overestimated, as expected. Conversely, if mutation rates specified in the analysis are larger than those used for the simulations, the age is underestimated. In both cases, this overestimation (or underestimation) becomes really important when $\mu_{sim} = 10^{-3}$ and $\mu_{ana} = 0$ (or $\mu_{sim} = 0$ and $\mu_{ana} = 10^{-3}$). Moreover, the bias is much more problematic with a tight map of markers ($\theta = 0.002$) with an overestimation (underestimation) of n_{gen} around 50% (30%), and p_{out-ci} values around 40%.

Table 3 presents the impact of the misspecification of allele frequencies for two different recombination fractions ($\theta = 0.01$ and $\theta = 0.002$). Three different true allele frequencies at the markers (0.1, 0.2, and 0.3) are considered and analysed under different assumed frequencies (0.01, 0.1, 0.2, and 0.3). When data are analysed with the correct allele frequencies (bold values of table 3) the proposed correction provides very consistent results except in the situation combining a large number of generations ($n_{gen} = 100$), a true allele frequency of 0.30, and $\theta = 0.01$. This phenomenon is amplified when data are analysed with misspecified allele frequencies. While the results remain quite satisfactory with $n_{gen} = 10$, a more important impact of allele frequencies is observed when $n_{gen} = 100$. In this latter case, the number of generations can be severely underestimated (or overestimated) when the allele frequency used for the analysis is underestimated (or overestimated) and $\theta = 0.01$. However, when

Table 1 Mean and median estimates over 5000 replicates of the number of generations for different true values, n_{gen} , and different sample sizes

Sample size	n_{gen}	θ^{\ddagger}	Estimates		
			mean (95% CI)*	median	p_{out-ci}^{\dagger}
2 individuals	10	0.0005	18.5 (4–74)	12	0.12
	20	0.0005	35.3 (7–143)	23	0.12
	100	0.0005	140.9 (35–511)	105	0.09
5 individuals	10	0.0005	11.4 (6–24)	10	0.06
	20	0.0005	22.9 (11–46)	21	0.06
	100	0.0005	113.7 (55–232)	103	0.07
10 individuals	10	0.01	10.7 (7–17)	10	0.04
	20	0.01	21.3 (13–34)	21	0.05
	100	0.01	100.9 (64–144)	98	0.02
30 individuals	10	0.01	10.2 (8–13)	10	0.02
	20	0.01	20.5 (16–26)	20	0.03
	100	0.01	101.8 (78–135)	100	0.05

*Empirical 95% confidence interval of the mean computed over the 5000 replicates.

[†]Proportion of replicates in which the 95% confidence interval of the estimate did not include the true value n_{gen} .

[‡]Recombination fractions between markers (20 markers were simulated on each side of the mutation).

Table 2 Influence of misspecifying mutation rates on the estimates of the number of generations according to the true number of generations (n_{gen}) and the recombination fraction between markers ($\theta = 0.01$ or 0.002)

Marker spacing	n_{gen}	Mutation rate for analysis	True mutation rate		
			$\mu_{sim} = 0$	$\mu_{sim} = 10^{-6}$	$\mu_{sim} = 10^{-3}$
$\theta = 0.01$	10	$\mu_{ana} = 0$	10 (0.04)*	10 (0.04)	11 (0.06)
		$\mu_{ana} = 10^{-6}$	10 (0.04)	10 (0.04)	11 (0.06)
		$\mu_{ana} = 10^{-3}$	9 (0.04)	9 (0.04)	10 (0.04)
	100	$\mu_{ana} = 0$	98 (0.02)	98 (0.01)	105 (0.01)
		$\mu_{ana} = 10^{-6}$	98 (0.02)	98 (0.01)	105 (0.01)
		$\mu_{ana} = 10^{-3}$	89 (0.05)	89 (0.04)	96 (0.02)
$\theta = 0.002$	10	$\mu_{ana} = 0$	10 (0.04)	10 (0.04)	15 (0.35)
		$\mu_{ana} = 10^{-6}$	10 (0.04)	10 (0.04)	15 (0.35)
		$\mu_{ana} = 10^{-3}$	7 (0.27)	7 (0.26)	10 (0.03)
	100	$\mu_{ana} = 0$	103 (0.06)	102 (0.06)	153 (0.45)
		$\mu_{ana} = 10^{-6}$	103 (0.06)	102 (0.06)	153 (0.45)
		$\mu_{ana} = 10^{-3}$	68 (0.37)	68 (0.38)	102 (0.06)

Results are presented for a sample size of 10 individuals. Values in bold indicate the situation in which the simulated data are analysed using the true mutation rate.

*Median estimate of the number of generations (p_{out-ci}) computed over 5000 replicates.

θ between markers is reduced to 0.002, the influence of the misspecification of allele frequencies on the estimates is much less important, arguing in favour of the use of tightly linked markers. The same observation is obtained in smaller samples (2 or 5 affected individuals) in which the use of low recombination fractions ($\theta = 0.0005$) between markers makes the analysis almost independent of allele frequencies (data not shown).

Triple-A study

Twelve markers were genotyped around the AAAS gene, six markers on each side, and the haplotypes of the nine patients at these markers are presented in table 4. For each of these patients, only one of the two haplotypes carrying the mutation was considered since the disorder is recessive and all families were inbred. The shared region is in bold italic in table 4 and all patients were homozygous in this region. Since

Table 3 Influence of misspecifying marker allele frequencies on the estimates of the number of generations according to the true number of generations (n_{gen}) and the recombination fraction between markers (θ)

Marker spacing	n_{gen}	Analysis frequency	True allele frequency		
			0.10	0.20	0.30
$\theta = 0.01$	10	0.01	10 (0.04)*	10 (0.03)	10 (0.03)
		0.10	10 (0.05)	10 (0.03)	10 (0.03)
		0.20	10 (0.04)	10 (0.04)	10 (0.03)
	100	0.30	11 (0.05)	10 (0.04)	10 (0.03)
		0.01	88 (0.04)	79 (0.13)	67 (0.33)
		0.10	98 (0.04)	85 (0.06)	73 (0.20)
$\theta = 0.002$	10	0.20	109 (0.07)	93 (0.05)	79 (0.11)
		0.30	116 (0.08)	101 (0.03)	85 (0.06)
		0.01	10 (0.04)	10 (0.04)	10 (0.04)
	100	0.10	10 (0.04)	10 (0.04)	10 (0.04)
		0.20	10 (0.04)	10 (0.04)	10 (0.04)
		0.30	10 (0.04)	10 (0.04)	10 (0.04)
100	0.01	100 (0.05)	96 (0.05)	93 (0.04)	
		0.10	102 (0.05)	99 (0.05)	95 (0.04)
		0.20	104 (0.07)	102 (0.05)	97 (0.05)
	0.30	107 (0.08)	104 (0.06)	100 (0.04)	

Results are presented for a sample size of 10 individuals. Values in bold indicate the situation in which the simulated data are analysed using the true marker allele frequency.

*Median estimate of the number of generations (p_{out-ci}) computed over 5000 replicates.

Table 4 Haplotypes encompassing AAAS in nine unrelated patients (P1–P9) sorted according to the length of sharing on the centromeric (left in table) side of the mutation

Patient	D12S 85	D12S 1701	D12S 368	D12S 96	D12S 359	D12S 1604	AAAS	D12S 1586	D12S 1707	D12S 1691	D12S 90	D12S 83	D12S 1702
P1	1	2	<i>1</i>	<i>1</i>	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>1</i>	<i>1</i>			
P2	5	2	<i>1</i>	<i>1</i>	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>1</i>	<i>6</i>	2	<i>1</i>	2
P3	4	2	<i>1</i>	<i>1</i>	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>1</i>	<i>3</i>			
P4		1	<i>1</i>	<i>1</i>	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>4</i>				
P5			2	<i>1</i>	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>2</i>				
P6			4	<i>1</i>	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>1</i>	<i>6</i>	2	<i>1</i>	3
P7			2	<i>1</i>	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>1</i>	<i>4</i>			
P8				<i>3</i>	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>2</i>				
P9				2	<i>1</i>	<i>5</i>	<i>IVS14+1G→A</i>	<i>3</i>	<i>3</i>				
Mb*	49.18	49.78	54.98	55.52	56.51	56.53	56.54	56.92	57.76	60.47	61.46	64.02	71.51
θ^{\dagger}	0.0508	0.0466	0.0108	0.0070	0.0002	0.0001		0.0026	0.0084	0.0271	0.0339	0.0516	0.1032
Freq [‡]		0.42	0.01	0.02	0.27	0.22		0.19	0.46	0.21	0.07	0.04	

The parts of the ancestral haplotype shared homozygously by all patients are indicated in bold italics and those shared by a subset of patients are indicated in italics.

*Physical distance in Mb obtained from www.ncbi.nlm.nih.gov. The genetic distance between D12S85 and D12S1702 is 15.9 cM.

[†] θ is the recombination rate between AAAS and the markers shown and was estimated using the correspondence 0.712 cM for 1 Mb over the whole region and Kosambi mapping function (see text).

[‡]Freq is the frequency of the marker allele shared by the patients (shown in italics) which was estimated from a sample of 30 unrelated subjects of north African origin.

the genetic and physical distances between D12S85 and D12S1702, the two extreme markers of the haplotype, are 15.9 cM (Genethon map¹⁶), and 22.33 Mb, respectively, as calculated by BLAST human genome searches, the correspondence between genetic and physical distances over the whole region was estimated to be 0.712 cM for 1 Mb. Recombination fractions between the different markers and AAAS were then computed from genetic distances using the Kosambi mapping function, and are shown in table 4.

The age of the most recent common ancestor of the IVS14+1G→A mutation was estimated to be 47 generations (95% confidence interval, 28 to 80) when the marker mutation rate was fixed at 0. Assuming that one generation is 25 years, this corresponds to 1175 years (95% confidence interval, 700 to 2000 years). When marker mutations were considered in the analysis, age estimates were reduced slightly to 46 generations (95% confidence interval, 28 to 78) when $\mu = 10^{-4}$, and to 39 (95% confidence interval, 24 to 67) generations when $\mu = 10^{-3}$.

To investigate whether this result can be affected if some subsets of patients are more related, we estimated the age of the mutation on subsamples of eight patients, removing each individual one after the other (assuming $\mu = 0$). These estimates varied from 41 (when individual P8 or P9 was excluded) to 88 generations (when individual P2 was excluded), and the 95% confidence interval of these estimates always included the initial estimate of 47 generations. Although this is not conclusive evidence against the existence of more related subsets of patients in our sample, this result indicates that this eventuality would have a small impact on the estimate. Moreover, no significant correlation ($p > 0.4$) was observed in the length of sharing between the two sides of the mutation as illustrated in table 4, where patient haplotypes have been sorted according to the length of sharing on the centromeric side of the disease causing mutation.

DISCUSSION

We have presented a simple likelihood based haplotype approach that estimates the age of the most recent common ancestor carrying the disease causing mutation, and not the age of the mutation itself, which would require assumptions about population genetic processes that are usually difficult to validate. As compared with other methods using haplotype information, our approach presents the advantage of being efficient with a very small number of affected individuals,

and is thus well suited to estimating the age of rare mutations. As expected, the method is particularly efficient when markers are very polymorphic and provides good estimates as long as shared alleles have frequencies less than 0.20 and these frequencies are correctly specified. An important result was the finding that the influence of allele frequencies on the results strongly decreases as marker distance decreases, suggesting that a helpful way to overcome the allele frequency issue is to refine the map surrounding the mutation. Another haplotype method that differs from ours in the way marker mutations are introduced and in the determination of the most plausible ancestral haplotype was also proposed to estimate the age of mutations in breast cancer genes.^{17–18} However, no simulations have been performed to investigate the reliability of this method in estimating the age of mutations and, in particular, rare mutations.

The method was used to date the IVS14+1G→A mutation in Triple-A syndrome using data from nine patients of north African origin. We found that the most recent common ancestor of these patients should have lived ≈ 1000 –1175 years (95% confidence interval, 600 to 2000) ago in a period of important migrations in north Africa from the Arabian peninsula.¹⁹ In this context, it is worth recalling that one of the first patients described with clinical features of Triple-A was from Saudi Arabia.²⁰ This analysis was performed assuming that all haplotypes observed in the sample diverged independently from a common ancestral haplotype (star genealogy) that carried the mutation. However, other kinds of intra-allelic genealogy can be assumed with subsets of patients sharing a more common ancestry.⁶ When prior information on the existence of subsets in the data is available (for example, different ethnic groups), an analysis stratified by subsets can be conducted. When no prior information is available, we are exploring strategies to identify subgroups based on haplotype sharing excess (for example, looking for correlation in the length of sharing between the two sides of the mutation which did not provide evidence for the existence of more related subsets of patients in the present Triple-A data).



Supplemental material is available online at <http://jmg.bmjournals.com/supplemental/>

Authors' affiliations

E Genin, Génétique Epidémiologique et Structure des Populations Humaines, INSERM U535, Hôpital Paul Brousse, BP 1000, 94 817 Villejuif Cedex, France

A Tullio-Pelet, F Begeot, S Lyonnet, Handicaps Génétiques de l'Enfant, INSERM U393, Hôpital Necker-Enfants Malades, 149 rue de Sèvres, 75015 Paris, France

L Abel, Génétique Humaine des Maladies Infectieuses, Université René Descartes INSERM U550, Faculté de Médecine Necker, 156 rue de Vaugirard, 75015 Paris, France

This work was supported by Fondation BNP-Paribas, Fondation pour la Recherche Médicale, Fondation Schlumberger, and Délégation à la Recherche Clinique (Grant CRC00114).

Conflicts of interest: none declared.

Correspondence to: Dr L Abel, Laboratoire de Génétique Humaine des Maladies Infectieuses, INSERM U550, Faculté de Médecine Necker, 156 rue de Vaugirard, 75015 Paris, France; abel@necker.fr

Revised version received 17 February 2004

Accepted for publication 17 February 2004

REFERENCES

- Allgrove J**, Clayden GS, Grant DB, Macaulay JC. Familial glucocorticoid deficiency with achalasia of the cardia and deficient tear production. *Lancet* 1978;*i*:1284–6.
- Weber A**, Wienker TF, Jung M, Easton D, Dean HJ, Heinrichs C, Reis A, Clark AJL. Linkage of the gene for the triple A syndrome to chromosome 12q13 near the type II keratin gene cluster. *Hum Mol Genet* 1996;*5*:2061–6.
- Hadj-Rabia S**, Salomon R, Pelet A, Penet C, Rotschild A, de Laët M-H, Chaouachi B, Hannaoui R, Bakiri F, Brauner R, Chaussain J-L, Munnich A, Lyonnet S. Linkage disequilibrium in inbred North African families allows fine genetic and physical mapping of triple A syndrome. *Eur J Hum Genet* 2000;*8*:613–20.
- Tullio-Pelet A**, Salomon R, Hadj-Rabia S, Mugnier C, de Laët MH, Chaouachi B, Bakiri F, Brottier P, Cattolico L, Penet C, Bégeot M, Naville D, Nicolino M, Chaussain JL, Weissenbach J, Munnich A, Lyonnet S. Mutant WD-repeat protein in triple-A syndrome. *Nat Genet* 2000;*26*:332–5.
- Rannala B**, Bertorelle G. Using linked markers to infer the age of a mutation. *Hum Mutat* 2001;*18*:87–100.
- Slatkin M**, Rannala B. Estimating allele age. *Annu Rev Genomics Hum Genet* 2000;*1*:225–49.
- Kimura M**, Ohta T. The age of a neutral mutant persisting in a finite population. *Genetics* 1973;*75*:199–212.
- Slatkin M**, Rannala B. Estimating the age of alleles by use of intraallelic variability. *Am J Hum Genet* 1997;*60*:447–58.
- Rannala B**, Reeve JP. High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence. *Am J Hum Genet* 2001;*69*:159–78.
- McPeck MS**, Strahs A. Assessment of linkage disequilibrium by the decay of haplotype sharing, with application to fine-scale genetic mapping. *Am J Hum Genet* 1999;*65*:858–75.
- Guo SW**, Xiong M. Estimating the age of mutant disease alleles based on linkage disequilibrium. *Hum Hered* 1997;*47*:315–37.
- Reeve JP**, Rannala B. DMLE+: Bayesian linkage disequilibrium gene mapping. *Bioinformatics* 2002;*18*:894–5.
- Piccolo F**, Jeanpierre M, Leturcq F, Dodé C, Azibi K, Toutain A, Merlini L, Jarre L, Navarro C, Krishnamoorthy R, Tomé FM, Urtizberea JA, Beckmann JS, Campbell KP, Kaplan J-C. A founder mutation in the γ -sarcoglycan gene of Gypsies possibly predating their migration out of India. *Hum Mol Genet* 1996;*5*:2019–22.
- Brinkmann B**, Klitschar M, Neuhuber F, Hühne J, Rolf B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 1998;*62*:1408–15.
- Picard C**, Fieschi C, Altare F, Al-Jumaah S, Al-Hajjar S, Feinberg J, Dupuis S, Soudais C, Al-Mohsen IZ, Génin E, Lamma S, Kumararatne DS, Leclerc T, Rafii A, Frayha H, Murugasu B, Wah LB, Sinniah R, Loubser M, Okamoto E, Al-Ghoniaim A, Tufenkeji H, Abel L, Casanova JL. Inherited interleukin-12 deficiency: *IL12B* genotype and clinical phenotype of 13 patients from six kindreds. *Am J Hum Genet* 2002;*70*:336–48.
- Dib C**, Faure S, Fizames C, Samson D, Drouot N, Vignal A, Millasseau P, Marc S, Hazan J, Seboun E, Lathrop M, Gyapay G, Morissette J, Weissenbach J. A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* 1996;*380*:152–4.
- Neuhausen SL**, Mazoyer S, Friedman L, Stratton M, Offit K, Caligo A, Tomlinson G, Cannon-Albright L, Bishop T, Kelsell D, Solomon E, Weber B, Couch F, Struwing J, Tonin P, Durocher F, Narod S, Skolnick MH, Lenoir G, Serova O, Ponder B, Stoppa-Lyonnet D, Easton D, King MC, Goldgar DE. Haplotype and phenotype analysis of six recurrent BRCA1 mutations in 61 families: results of an international study. *Am J Hum Genet* 1996;*58*:271–80.
- Neuhausen SL**, Godwin AK, Gershoni-Baruch R, Schubert E, Garber J, Stoppa-Lyonnet D, Olah E, Csokay B, Serova O, Lalloo F, Osorio A, Stratton M, Offit K, Boyd J, Caligo MA, Scott RJ, Schofield A, Teugels E, Schwab M, Cannon-Albright L, Bishop T, Easton D, Benitez J, King MC, Goldgar D. Haplotype and phenotype analysis of nine recurrent BRCA2 mutations in 111 families: results of an international study. *Am J Hum Genet* 1998;*62*:1381–8.
- Nebel A**, Landau-Tasseran E, Filon D, Oppenheim A, Faerman M. Genetic evidence for the expansion of Arabian tribes into the Southern Levant and North Africa. *Am J Hum Genet* 2002;*70*:1594–6.
- Lanes R**, Plotnick LP, Bynum TE, Lee PA, Casella JF, Fox CE, Kowarski AA, Migeon CJ. Glucocorticoid and partial mineralocorticoid deficiency associated with achalasia. *J Clin Endocrinol Metab* 1980;*50*:268–70.



Estimating the age of rare disease mutations: the example of Triple-A syndrome

E Genin, A Tullio-Pelet, F Begeot, S Lyonnet and L Abel

J Med Genet 2004 41: 445-449
doi: 10.1136/jmg.2003.017962

Updated information and services can be found at:
<http://jmg.bmj.com/content/41/6/445>

	<i>These include:</i>
References	This article cites 20 articles, 4 of which you can access for free at: http://jmg.bmj.com/content/41/6/445#BIBL
Email alerting service	Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections	Articles on similar topics can be found in the following collections Molecular genetics (1254) Genetic screening / counselling (886) Adrenal disorders (7)
--------------------------	---

Notes

To request permissions go to:
<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:
<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:
<http://group.bmj.com/subscribe/>