MICROSATELLITES: SIMPLE SEQUENCES WITH COMPLEX EVOLUTION

Hans Ellegren

Few genetic markers, if any, have found such widespread use as microsatellites, or simple/short tandem repeats. Features such as hypervariability and ubiquitous occurrence explain their usefulness, but these features also pose several questions. For example, why are microsatellites so abundant, why are they so polymorphic and by what mechanism do they mutate? Most importantly, what governs the intricate balance between the frequent genesis and expansion of simple repetitive arrays, and the fact that microsatellite repeats rarely reach appreciable lengths? In other words, how do microsatellites evolve?

HETEROZYGOSITY
The proportion of individuals in a population that carry two different alleles at a locus.

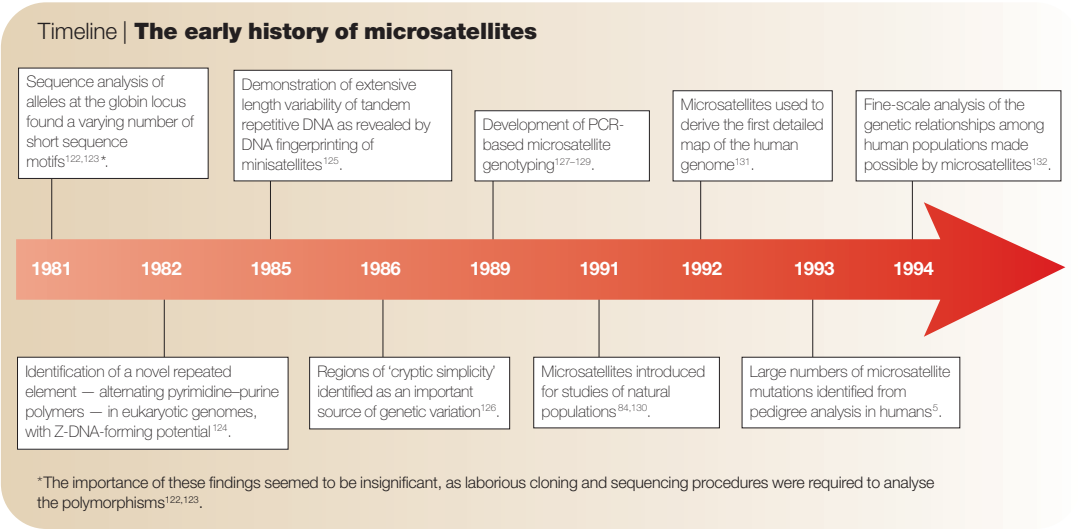
GENE FLOW
The transfer of alleles within and between populations that arises from migration and dispersal.

Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, Norbyvägen 18D, SE-752 36 Uppsala, Sweden.
e-mail: Hans.Ellegren@ebc.uu.se
doi:10.1038/nrg1348

Assuming chance association of nucleotides, the probability of finding the sequence CACACACACACACACACACA more than once in the human genome is negligible. However, perfect or near-perfect tandem iterations of short sequence motifs of this kind are extremely common in eukaryotic genomes and, in the case of the human genome, they are found at hundreds of thousands of places along chromosomes[1]. This particular genomic feature is not restricted to $(CA)_n$ repeats — every possible motif of mono-, di, tri- and tetranucleotide repeats is vastly overrepresented in the genome. Ever since their discovery in the early 1980s, the ubiquitous occurrence of microsatellites — also referred to as short tandem repeats (STRs) or simple sequence repeats (SSRs) — has puzzled geneticists. Why are they so common? Do they fulfill some function or are they simply junk DNA sequences that should perhaps be viewed as 'selfish DNA'[2,3]? Addressing these questions is important if we wish to understand how genomes are organized and why most genomes are filled with sequences other than genes.

Microsatellites are among the most variable types of DNA sequence in the genome[4]. In contrast to unique DNA, microsatellite polymorphisms derive mainly from variability in length rather than in the primary sequence. Moreover, genetic variation at many microsatellite loci is characterized by high HETEROZYGOSITY and the presence of multiple alleles, which is in sharp contrast to unique DNA. With the advent of PCR in the late 1980s, the analysis and genotyping of microsatellite polymorphisms became straightforward (see TIMELINE). Microsatellites quickly became the marker of choice in genome mapping, and subsequently also in population genetics studies and related areas.

For a neutral marker, the degree of polymorphism is proportional to the underlying rate of mutation. Given the extensive polymorphism of microsatellites, it follows that mutations must occur frequently — an assumption that is supported by direct observations[5]. The rate and direction of mutations constitute two basic factors in the estimation of genetic distance on the basis of microsatellite data. By applying theoretical models of microsatellite evolution to empirical data, population geneticists attempt to, for example, determine how long ago two populations diverged, or measure the amount of GENE FLOW between populations. However, despite the extensive use of microsatellite markers over the past 15 years, it is clear that many theoretical models fail to accurately explain allele frequency distributions in natural populations. Importantly, it seems that microsatellite evolution is a far more complex process than was previously thought. A deeper understanding of the evolutionary and mutational properties of microsatellites is therefore needed, not



only to understand how the genome is organized, but also to correctly interpret and use microsatellite data in population genetics studies.

Recent new information provides clues to the mystery of microsatellite repeats. First, whole-genome sequence data provide an unbiased picture of the occurrence and genomic distribution of repetitive elements. Second, large-scale pedigree analysis in different organisms gives direct insight into the characteristics of *de novo* mutation events. Third, molecular studies of the DNA replication machinery show what might go wrong during microsatellite replication. Here, I review these new findings and summarize our current knowledge about microsatellite evolution. Emerging from the new data is the picture of a heterogeneous mutation process, showing distinct differences in rates and patterns of mutation among loci and species.

The genome biology of microsatellites

What is a microsatellite? Genomes are scattered with simple repeats. Tandem repeats occur in the form of iterations of repeat units of almost anything from a single base pair to thousands of base pairs. Mono-, di-, tri- and tetranucleotide repeats are the main types of microsatellite, but repeats of five (penta-) or six (hexa-) nucleotides are usually classified as microsatellites as well. Repeats of longer units form minisatellites or, in the extreme case, satellite DNA. The term satellite DNA originates from the observation in the 1960s of a fraction of sheared DNA that showed a distinct buoyant density, detectable as a 'satellite peak' in DENSITY GRADIENT CENTRIFUGATION, and that was subsequently identified as large centromeric tandem repeats. When shorter (10–30-bp) tandem repeats were later identified, they came to be known as minisatellites. Finally, with the discovery of tandem iterations of simple sequence motifs, the term microsatellites was coined. The difference between the terms micro- and minisatellites might not be obvious *per se*, but it is motivated by the difference in the mutational mechanisms of repeats of just a few nucleotides and of ten or more (see below).

It is more complicated to define the minimum number of iterations needed for a repetitive sequence to be referred to as a microsatellite. For instance, the sequence CACA occurs frequently in the human genome: should it be seen as $(CA)_2$ microsatellites or just as unique sequence? In practice, the threshold that is used when describing the occurrence of a microsatellite in a genomic sample data set must be specified. Unfortunately, no real consensus has been reached on this matter; whereas some use a minimum number of base pairs, others use a minimum number of repeat units, and in both cases, the numbers have varied. The issue is further complicated by the lack of agreement on how much degeneracy should be accepted for characterizing a slightly imperfect tandem repetitive sequence as a microsatellite. Mismatch considerations are particularly important when using algorithms (such as RepeatMasker, Sputnik and Tandem Repeats Finder; see online links box and BOX 1) to search large genomic sequences for repeats.

It is appropriate to further classify microsatellites according to their association with coding sequence as this is related to the mutational and selective forces that operate on different types of repeat. The bulk of simple repeats are embedded in non-coding DNA, either in the intergenic sequence or in the introns. Microsatellites that are used as genetic markers are usually of this type and are generally assumed to evolve neutrally. Their frequency and distribution should therefore reflect the underlying mutation process. In coding DNA, selection against frameshift mutations effectively hinders the expansion of everything other than trinucleotide repeats[6], for which there might be further length constraints related to protein function[7]. Trinucleotide repeats associated with human disease comprise a special class of microsatellites in coding DNA. These loci undergo extensive repeat expansions, the mutational mechanism of which is thought to differ from that of most microsatellites in the genome. For instance, the establishment of hairpin structures with a relatively high amount of base-pair complementarities might stabilize loops that are generated during replication slippage.

DENSITY GRADIENT CENTRIFUGATION
Separation of biomolecules on the basis of their density.

Details of the evolution of expanded trinucleotide repeats have been described elsewhere and will not be considered further here.

Microsatellite distribution. The initial analysis of the draft sequence of the human genome concluded that microsatellites account for 3% of the genome[1]. There are more than one million microsatellite loci in the human genome, although the exact number greatly depends on the parameters of the search algorithm (for example, gap and mismatch penalties). This number also includes an appreciable proportion of interrupted microsatellites and many that are probably monomorphic. Dinucleotide repeats dominate, followed by mono- and tetranucleotide repeats, and trinucleotide repeats are least dominant. Again, however, it is a matter of how microsatellites are defined. Among repeats that are at least 12 bp long, mononucleotide repeats outnumber dinucleotide repeats; the reverse situation is not valid until a higher threshold is used. Among dinucleotides, $(CA)_n$ repeats are most frequent, followed by $(AT)_n$, $(GA)_n$ and $(GC)_n$, the last type of repeat being rare. Note that there are only four possible types of dinucleotide repeat, because CA = AC = GT = TG, GA = AG = CT = TC, AT = TA, and GC = CG.

Data from the mouse genome have confirmed the abundance of microsatellites but have also revealed impressive differences[8]. If identical search criteria are used, the mouse genome proves to be repeat-rich with two–threefold more microsatellites than humans. Moreover, microsatellites are longer in mice than in humans, and the same holds true for the rat–human comparison[9]. Preliminary data from other mammalian genomes indicate that rodent genomes have particularly high microsatellite numbers. This might be a general phenomenon — that microsatellite occurrence differs between related species. In fact, differences might even occur between such closely related species as humans and chimpanzees[10], and within the genus *Drosophila*[11,12].

Microsatellite density tends to positively correlate with genome size[13–15]. Among fully sequenced eukaryotic genomes, microsatellite density is highest in mammals. However, in plants, microsatellite frequency is negatively correlated with genome size[16]. This has been attributed to the fact that microsatellites are underrepresented in the repetitive parts of the plant genome that are involved in genome expansion, such as the long terminal repeats of RETROTRANSPOSONS[16]. Another peculiar feature of most plant genomes is that $(AT)_n$ is the most common motif among dinucleotides[17]. Assuming that, on a genomic scale, microsatellite sequences are at equilibrium, the contrasting distributions of microsatellite motifs in different genomes strongly indicate that there is interspecific variation in the mechanisms of mutation or repair of specific motifs. Alternatively, there might be variation in the selective constraints that are associated with different microsatellite motifs.

Are microsatellites equally common everywhere in the genome? There seem to be no distinct differences in density between intergenic regions and introns[14]. Base composition influences microsatellite density, which is consistent with their neutral origin and random generation by mutation[18]. There is, however, evidence for regional variation in microsatellite frequency that cannot be explained by base composition[18], and, in the human and mouse genomes, microsatellite density is nearly twofold higher near the ends of chromosome arms[8]. What accounts for this heterogeneity remains to be explained. In several species, the density and/or the length distribution of microsatellites on the X chromosome differs from that on the autosomes[8,18]. It might be a result of factors such as sex differences in the mutation rate, differences in EFFECTIVE POPULATION SIZE between the X chromosome and autosomes, and the efficiency of selection on hemizygous chromosomes.

Microsatellites are also frequently found in the proximity of interspersed repetitive elements such as short interspersed repeats (SINEs) and long interspersed elements (LINEs). For example, human *Alu* repeats often have a microsatellite-like structure at their 3′ ends[19] that might arise from the introduction of poly(A) tails of reversed transcribed messages when element insertion takes place. This is consistent with the observation that mononucleotide arrays $(A)_n$ and other types of A-rich microsatellite dominate at these sites. Other examples include the intimate association of microsatellites with

Box 1 | Informatics approaches to finding microsatellites in a genomic data set

Sputnik

Sputnik uses a recursive algorithm to search for repeats of two–five nucleotides in sequence files in FASTA format. Insertions, mismatches and deletions are tolerated but affect the overall score. If the score falls below a cutoff threshold, the search is abandoned and begun again at the next nucleotide. Sputnik does not compute an entire identity matrix first and then pick the best of the hits; instead, it starts at the beginning and compares the patterns until the score falls below a cutoff threshold.

RepeatMasker

RepeatMasker does not use a recursive algorithm. It scans for di- to pentameric repeats and simple repeats that are shorter than 20 bp, and those with >10% divergence from a perfect repeat are ignored.

Tandem Repeats Finder

This program works without the need to specify either the pattern or the pattern size. It models tandem repeats according to the percentage identity and frequency of indels (insertions/deletions) between adjacent pattern copies and uses statistically-based recognition criteria. The program can return a copy of the original sequence with the tandem repeats masked out.

RETROTRANSPOSONS
Mobile elements that spread in the genome through an RNA intermediate.

EFFECTIVE POPULATION SIZE
The theoretical size of an idealized population that has the same magnitude of random genetic drift as the actual population.

retrotransposon-like elements in barley[20], and $(AT)_n$ microsatellites that are frequently found to be juxtaposed with miniature inverted repeat-transposable elements (Micropon-4) in rice[21]. In some of these cases, microsatellites have evolved from internal A-rich structures, and it is also possible that insertion of an interspersed element might in itself be favoured at sites with a pre-existing microsatellite[22].

Microsatellites are present in low numbers in prokaryotes. This is particularly true for longer repeats, for which the numbers are lower than would be expected on the basis of nucleotide composition[23], in sharp contrast to the situation in eukaryotic genomes. Even short prokaryotic microsatellites might still vary in length[24]. Unusually long microsatellites are sometimes associated with virulence factors, in which case, they act as translation and transcriptional 'switches'; therefore, their presence is maintained by positive selection[25].

The mutation process

Mutation models. A mutation model of microsatellite evolution is needed if allele frequency data from two groups of individuals (for example, populations or species) are to be used for estimating the genetic distance between them. A wide range of models of the evolutionary dynamics of microsatellites has been presented, most of which derive from the stepwise mutation model (SMM)[26] (FIG. 1). Adopted for microsatellites, the original SMM postulates that a mutation alters the length of a repetitive array through the addition or removal of one repeat unit at a fixed rate (a symmetric forward–backward random walk that is independent of repeat length)[27–30]. However, it soon became apparent that a simple SMM does not lead to stationary microsatellite-length distributions[31]. For example, the fact that microsatellites seem to show an upper size limit is incompatible with the SMM. Extensions of the SMM have therefore introduced an upper limit on allele sizes[32–35], or a mutational bias such that large alleles mutate preferentially to alleles of smaller sizes[36–38]. Other approaches have involved more complex stepwise models[39–42]. The parameters of the mathematical models are tested against measures of variability (heterozygosity, variance of repeat counts, SKEWNESS) that are observed within populations, and, more recently, against microsatellite distributions in genomic data sets.

An attractive model of microsatellite evolution holds that a genome-wide distribution of microsatellite repeat length that is at equilibrium results from a balance between length and point mutations[43,44]. According to this model, two opposing mutational forces operate on microsatellite sequences. Length mutations, the rate of which increases with increasing repeat count, favour loci to attain arbitrarily high values, whereas point mutations break long repeat arrays into smaller units. At equilibrium, there will be a steady-state distribution of repeat lengths governed by the rate of length mutation and the rate of point mutation. This model, or derivatives thereof, has been well received in recent years because it can explain differences in microsatellite

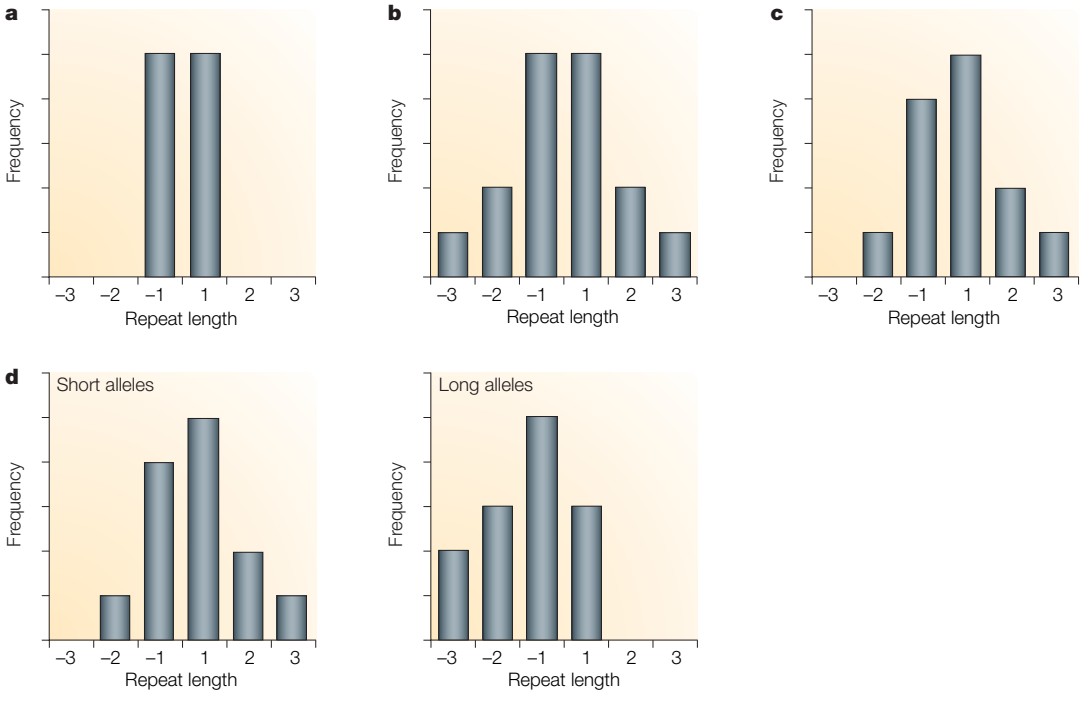


Figure 1 | **Microsatellite mutation models.** The magnitude and direction of mutation events according to different forms of the stepwise mutation model. Mutations are indicated by the change in the number of repeat units; for example, +1 is an expansion of one repeat unit. **a** | A simple model that only involves one-step changes. **b** | A model that involves multi-step changes. **c** | A model with directionality in favour of repeat expansions. **d** | A length-dependent model in which short alleles tend to increase in size, whereas long alleles show a bias towards contraction.

SKEWNESS
Deviation from the normal distribution.

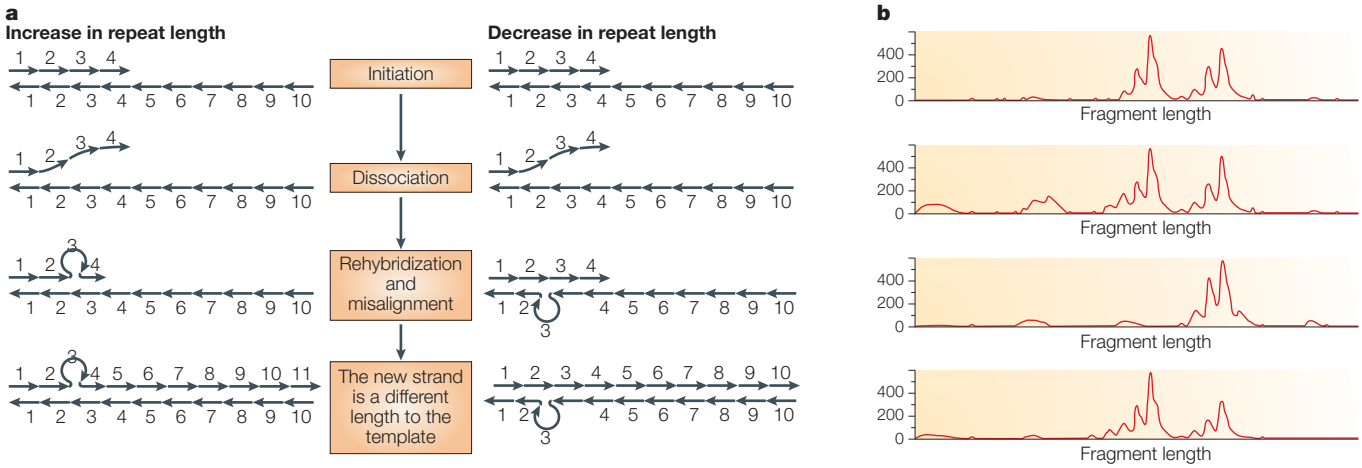


Figure 2 | **Replication slippage. a** | After the replication of a repeat tract has been initiated, the two strands might dissociate. If the nascent strand then realigns out of register, continued replication will lead to a different length from the template strand. If misalignment introduced a loop on the nascent strand, the end result would be an increase in repeat length. A loop that is formed in the template strand leads to a decrease in repeat length. **b** | Replication slippage also occurs during *in vitro* amplification of microsatellites, in this case mainly in the form of repeat contractions. These events can be recognized as minor peaks — known as stutter bands — that differ from each main product by multiples of the repeat unit length.

distribution among species and provides an elegant solution to the problem of why microsatellites do not expand into enormous arrays. However, even with this model, evolutionary dating of divergence times is not necessarily trivial[45].

Mutation mechanism. Length changes in microsatellite DNA are generally thought to arise from replication slippage — that is, transient dissociation of the replicating DNA strands followed by misaligned reassociation[46] (FIG. 2a). When the nascent strand realigns out of register, renewed replication will lead to the insertion or deletion of repeat units relative to the template strand. Most of these primary mutations are corrected by the MISMATCH-REPAIR SYSTEM, and only the small fraction that was not repaired ends up as microsatellite mutation events[47]. *In vitro* experiments that use purified eukaryotic or prokaryotic enzymes confirm that DNA polymerase is the only enzymatic activity needed for slippage[48]. Slippage involves DNA polymerase pausing, during which the polymerase dissociates from the DNA. On dissociation, only the terminal portion of the newly synthesized strand separates from the template and subsequently anneals to another repeat unit[49].

Replication slippage also occurs during PCR amplification of microsatellite sequences *in vitro* (FIG. 2b). A characteristic feature of such amplifications is the presence of 'stutter bands' — that is, minor products that differ in size from the main product by multiples of the length of the repeat unit[50,51]. Quantitative experiments show that the *Taq* polymerase slippage rate increases with the number of repeat units and is inversely correlated with repeat unit length[52]. PCR-induced stutter bands have been observed by many microsatellite users; tetranucleotide repeat markers typically give fewer stutter bands than dinucleotide and, in particular, mononucleotide repeats. Stutter bands generally appear as products that are shorter than the size of the allele being amplified. The *in vitro* rate of contraction mutation caused by *Taq* polymerase must therefore be much higher than the rate of expansions[52]. For this reason, PCR slippage probably cannot be used to gain insight into the mutational dynamics of microsatellites *in vivo*.

Recombination-like processes that involve unequal crossover or GENE CONVERSION introduce mutations in the larger minisatellite sequences[53]. There is little evidence that recombination would also contribute to microsatellite mutations. Genomic microsatellite distributions are associated with sites of recombination[54], most probably as a consequence of repetitive sequences being involved in recombination rather than being a consequence of it[55]. Moreover, most tests for a correlation between recombination rate and microsatellite density or mutability have failed to demonstrate such an effect[18,56]. Furthermore, there is no evidence of any systematic differences in the rates and patterns of microsatellite mutations between autosomal and Y-linked markers; the fact that Y-chromosome sequences are not involved in meiotic recombination therefore does not influence the mutation process[57,58]. Neither do the observations of similar patterns of microsatellite mutations in somatic cells[59] and in germ cells support a role for recombination.

Character of observed mutations. Although improved mutation models are now available, it can be difficult to assess to what extent they reflect true evolutionary processes. Fortunately, the high rate of mutation at microsatellite loci makes it possible to observe mutation events directly. Specifically, pedigree analysis offers a means for mutation detection (BOX 2), and data on *de novo* mutations have now been reported for a range of loci and organisms[5,60–64]. The general pattern that emerges is compatible with replication slippage in which

MISMATCH-REPAIR SYSTEM
An enzymatic system for the correction of errors that are introduced during DNA replication or recombination when an incorrect base is incorporated into the daughter strand, or when small insertion–deletion loops are being formed.

GENE CONVERSION
A meiotic process of directed change in which one allele directs the conversion of a partner allele to its own form.

new variants differ from their progenitor alleles by integral numbers of repeats. The fact that mutations sometimes involve more than one repeat unit means that the single-step mutation model is not valid in most cases.

One important conclusion from observations of spontaneous mutations is that the mutation process seems to be heterogeneous with respect to loci, repeat types and organisms. For instance, most human studies find that <15% of mutation events are multi-step changes[5,58,65–68]. However, the three largest human studies of this kind that have so far been presented reveal contradictory results. In agreement with other studies, Ellegren[69] and Xu *et al.*[70] found 11–14% multi-step mutations among 102 and 236 mutation events, respectively. By contrast, Huang *et al.*[56], in an analysis of 97 mutation events, reported 63% multi-step changes. What accounts for this discrepancy remains unclear, but it might indicate that some loci are more prone to large changes than others. Analyses of individual loci in other organisms have revealed highly variable proportions of multi-step changes, in the range of 5–75% (REFS 62,63,71–73). A more extensive screening of zebrafish markers found 68% multi-step changes[61].

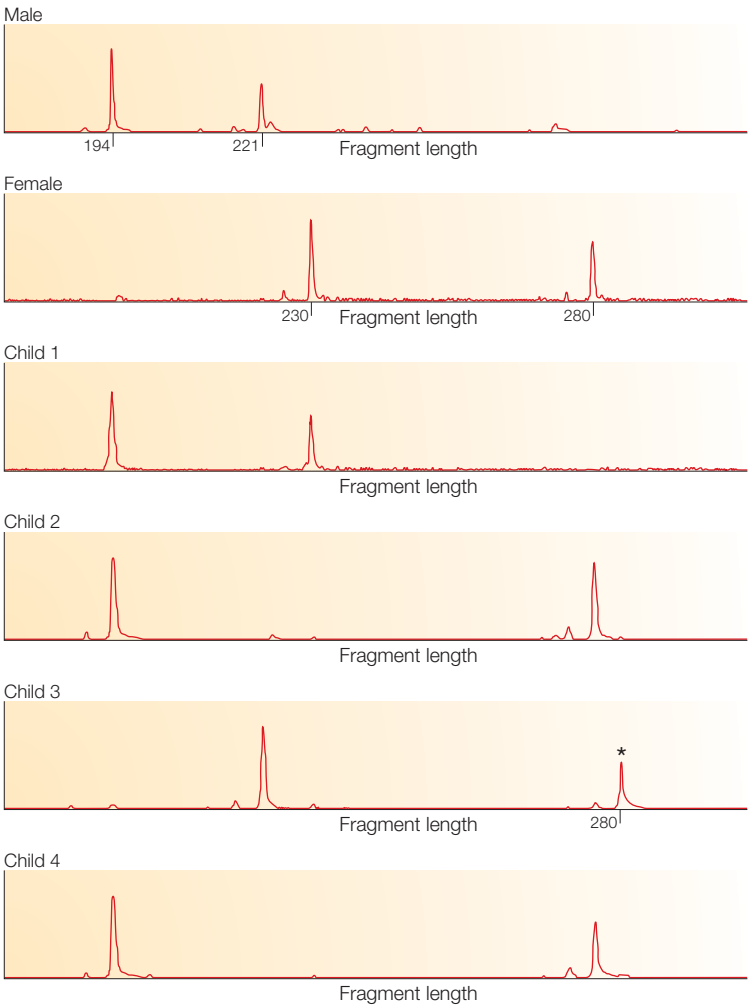
Heterogeneity is also seen in the propensity of mutation events to lead to different forms of alterations in microsatellite size. Directionality in the mutation process in favour of gains over losses has been observed for many human markers[39,58,65,66,68,69] and for bird microsatellites[71,72]. However, Xu *et al.*[70] found no such bias, and Huang *et al.*[56] found only a modest excess of contractions in their studies of human microsatellites. Whatever the cause of this heterogeneity, it will be interesting to see whether directionality is related to microsatellite length at the level of individual loci. Everything else being equal, we should expect loci that have a tendency to expand by mutation to grow more often than those that tend to contract. To add to the complexity, several studies have found evidence of a negative correlation between direction/magnitude of mutation and allele size[56,63,64,69,70,72,74] — that is, long alleles being biased towards contraction. If generally true, this would offer a mechanistic explanation for the stationary genomic length distributions seen at microsatellite loci. Interestingly, mutations from three bacterial species show a downward bias, which can perhaps account for the rarity of microsatellites in prokaryotic genomes[75].

Box 2 | **Using pedigrees to detect microsatellite mutations**

The most straightforward approach for the study of microsatellite mutations is direct observations of allele transmissions in parent–child pairs (see figure). A mutant allele (asterisk) is identified as being incompatible with Mendelian inheritance because it is different in size from parents' alleles. Data from other markers are needed to confirm that non-congruence between parental and offspring genotypes is not a result of incorrect parentage[133]. One limitation of this approach is that detectable mutations are restricted to cases in which child genotypes cannot be generated by transmission from parents' genotypes. A mutation to a character state that is already present in one of the parents can therefore remain undetected. However, likelihood-based estimates of mutation rates that take such events into account have been developed[134].

A related concern derives from the fact that even when the mutant allele is different from parents' alleles, it is not evident from which progenitor allele it originates. The standard assumption made is that the smallest mutational change is the most probable. However, this introduces circularity when such data are used to support the stepwise mutation model. Fortunately, simulations show that this assumption is valid in most cases, at least when compared with the alternative of random assignment to the progenitor allele[73].

An alternative approach for the identification of germline mutations is offered by the analysis of single sperm, or small pools of sperm[67,97]. This gives virtually unlimited access to gametes from specific males to screen, allowing individual-specific estimates of the mutation rate to be made. However, a disadvantage with this approach is that it is technically demanding to amplify such small amounts of DNA, and that it therefore requires most stringent laboratory routines to avoid contamination. On the other hand, as it does not require access to pedigrees, it might be useful for applications such as genetic toxicology screenings.



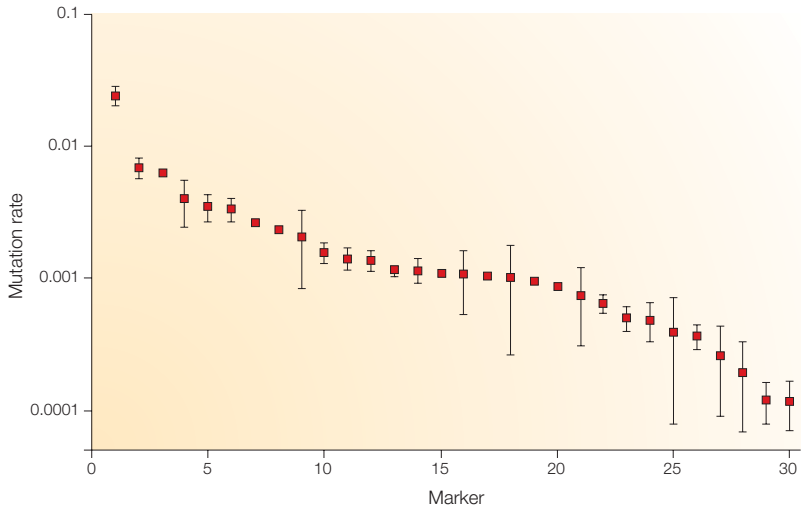


Figure 3 | **Microsatellite mutation rates in the human genome.** Observed sex-average mutation rates (log scale, with 95% confidence intervals) for human microsatellites obtained from pedigree analysis. Markers are given in the following order: 1, *D10S1214*; 2, *D12S1090*; 3, *ACTBP2*; 4, *D19S253*; 5, *D9S302*; 6, *D3S1744*; 7, *FGA*; 8, *HUMvWA31*; 9, *D22S683*; 10, *D18S51*; 11, *D8S1179*; 12, *D21S11*; 13, *CYP19*; 14, *D3S1358*; 15, *HUMCSF1P0*; 16, *D18S849*; 17, *D13S317*; 18, *D17S5*; 19, *D5S818*; 20, *D7S820*; 21, *Penta E*; 22, *D16S539*; 23, *HUMFESFPS*; 24, *HUMF13A01*; 25, *HUMLIPOL*; 26, *D1S80*; 27, *HUMF13B*; 28, *D2S1338*; 29, *HUMTH01*; and 30, *HUMTPOX*. Data from The Annual Report Summary for 2000 from the US Parentage Testing Standards Committee.

Sequence data. Another empirical approach to studying microsatellite evolution involves characterizing the sequence structure of alleles within species, or comparing the sequence of orthologous loci in different species. Using this approach, the effect of mutations accumulated over evolutionary timescales can be readily studied, although it might be difficult to determine the precise order and character of individual mutation events if they have occurred at high rates during the time period being surveyed. Note that in contrast to base substitutions, for which an infinite allele model is applicable, microsatellite alleles are often identical in state (structure) but not by descent[76]. For example, two chromosomes that are drawn from a population with the sequence $(GT)_{17}$ at a particular microsatellite locus might have reached this state through mutations from $(GT)_{16}$ or $(GT)_{18}$ alleles.

Nevertheless, sequence analysis has shed light on the genesis of new microsatellite loci, particularly in cases in which changes in orthologous microsatellite sequences can be mapped on a phylogenetic tree[77–81]. Such studies confirm that short repetitive sequences — with as few as two or three repeat units — are the starting point for subsequent microsatellite expansion. These primary repeats can arise from normal base substitutions, such as an A–G transition in GTATGT to GTGTGT. In addition, a significant proportion of new two-repeat loci arise from insertion mutations that are duplications of adjacent sequence[82]. Interestingly, this would be compatible with a recent model proposed by Dieringer and Schlotterer[42], which indicates that a length-independent mutation process operates on short microsatellites.

Not only the birth but also the death of microsatellites can be captured by sequence comparisons[83]. Such studies have supported the idea that, in the long run, point mutations break up perfect repeats and reduce the mutation rates of microsatellite loci. Clearly, long microsatellite alleles do not persist indefinitely. However, microsatellite evolution is a dynamic process; therefore, repeats might shrink as well as expand over evolutionary timescales. In fact, the removal of microsatellite interruptions by replication slippage[81] means that point mutations in microsatellite arrays do not necessarily lead to decay but might represent only a transition state during the evolution of microsatellites. This feature should be incorporated into the model of Kruglyak *et al.*[44] and others[43], which indicates that point mutations usually destroy perfect repeats.

Intraspecific comparisons that reveal the sequence structure of individual alleles provide further evidence for the complexity of the mutation process[36,77,79,84,85]. There are numerous loci in a range of organisms at which alleles differ not only in repeat length but also in repeat structure. Perhaps an extreme example of allele structure comes from 3 sequenced alleles at a bird tetranucleotide repeat locus: allele 1, $(AAAG)_{12}$; allele 2, $(AAAG)_{22}A(AAAG)_{12}$; and allele 3, $(AAAGAGAG)_6(A)_4(AG)_3(AAAG)_3(AG)_9AA(AG)_3(AAAG)_2(AG)_2(AAAG)_2(AGAGAAAG)_{15}(AAAG)_{24}$ (REF. 80).

Mutation-rate variation

There is no uniform microsatellite mutation rate; the rates differ among loci and among alleles, and, perhaps as a consequence, also among species[86]. The single most important factor to affect mutation rate that has so far been discovered is microsatellite length — mutation rate increases with an increasing number of repeat units. Intuitively, this seems understandable — more repeat units give more opportunities for replication slippage. A length-dependent mutation rate explains part of the mutation-rate variation at several scales. The low mutation rate in *Drosophila melanogaster* is compatible with microsatellites being much shorter in flies than in vertebrates[87]. Within species, measures of repeat lengths correlate with mutability[88]. Furthermore, a positive correlation between allele size and mutation rate has been seen in many organisms[60,66,72–74]. The precise character of the relationship between repeat count and mutation rate is less clear. Some studies have found a linear relationship[89], whereas some recent data indicate a power or exponential relationship between size and rate[10,88].

But repeat length is not the sole cause of microsatellite mutation-rate variation. As can be seen in FIG. 3, the mutation rate of individual loci in a selected set of human markers varies within two orders of magnitude, which cannot be attributed to repeat length. Similar observations are made in other organisms[62,63,72,90]. One important consequence of this variation is that the mean mutation rate of a set of markers will vary a lot depending on which particular markers are used. Moreover, as most markers that are used in genetic studies are selected on the basis of being (highly) polymorphic and,

given the expected relationship between polymorphism and mutability, the observed rates might not provide a representative picture for the genome as a whole.

What else matters? One possibility is that sequences that flank the microsatellite affect the mutation rate[91,92]. That inherent characteristics of individual loci are involved is indicated by covariation in levels of variability at orthologous loci in related species[93]. However, in addition to a flanking-sequence effect, this observation could also be compatible with an effect of, for example, TRANSCRIPTION-COUPLED REPAIR[18,94], chromatin structure, regional sequence context and local point-mutation-rate variation. As for the last possibility, an extension of the balance model of microsatellite evolution states that not only will the equilibrium length distribution of simple repeats be dependent on the species-specific rate of point mutation, but the length of individual repeat loci will also depend on the local point-mutation rate, for which there is evidence for significant heterogeneity within genomes[95]. In a study of orthologous microsatellite loci in the mouse and rat, a negative correlation between microsatellite length and substitution rate in nearby flanking sequence was found[96]. This would indicate that when point mutations occur at high frequency, they seem to hinder further microsatellite expansion (mutability) by introducing interruptions or imperfections in the repeat array.

Assuming a replication origin of microsatellite mutations, we should expect the mutation rate to correlate with the number of germline cell divisions. By extension, mutations should be more frequent in males than in females, and in older males than in younger males. Although seemingly reasonable, these predictions are only partly supported by empirical data. Two of the large studies of human mutations find three–four times more mutations in men[69,70], which is close to recent data on the male-to-female mutation-rate ratio for point mutations. However, Huang *et al.* saw no sex bias in microsatellite mutation rate in their study. Moreover, significant variation in the mutational sex-bias has been documented for swallow microsatellites, with at least one locus showing a male-biased rate, whereas others have female-biased rates[60,72]. Female-biased rates for individual loci have been reported for other organisms as well[73,90]. Attempts to correlate human microsatellite mutation rates with the father's age have either failed to find such an effect[58,68,97] or only found a small effect[66].

Experimental approaches

The empirical data on microsatellite mutations described above all refer to spontaneous events observed after germline transmissions. An experimental approach to the study of microsatellite evolution is offered by the analysis of instability of artificial plasmid-borne microsatellite sequences introduced into bacterial or eukaryotic cells. By constructing plasmids with repeats that are associated with a resistance or a reporter gene, length mutations that disrupt or restore the reading frame of that gene can easily be monitored. Mutation profiles have in this way been particularly well characterized in the yeast *Saccharomyces cerevisiae*, in which repeats that are integrated into chromosomes have also been studied.

These studies confirm several observations from germline transmissions. Mismatch repair is identified as crucial to microsatellite stability as mutation rates in prokaryotic and eukaryotic cells that are deficient in mismatch repair are increased by several orders of magnitude compared with wild-type cells[46,47,98,99]. Mutations in genes that encode proof-reading exonucleases and some DNA polymerases have also been implicated in repeat instability, although they have a more modest effect compared with mismatch-repair deficiency[100–102]. In all systems, mutation rate increases with repeat length[46,103,104], but interruptions stabilize repeat tracts[105,106]. The orientation of repeats — that is, whether a particular motif is on the coding or the complementary strand — does not seem to affect the mutation rate[98], with the exception of long trinucleotide repeat arrays[107]. Observations of the destabilization of microsatellites by elevated levels of transcription would support a role for transcription-coupled repair[103].

The effect of sequence composition on the relative instability of repeats is less clear. A study in *Escherichia coli* found no significant difference in the mutability of CA- and GA-repeats of similar length[108], in contrast to observations in human cells[59] (see also REF. 92). Conflicting observations are also made with respect to the effect of the length of the repeat unit, as seen, for example, in di- versus tetranucleotide repeats[99,109]. However, in *E. coli*[110] and in human[111] and yeast[112] cells, the mutability of G-mononucleotide repeats is higher than that of A-repeats of the same length, potentially owing to stronger stacking interactions among Gs or Cs than among As and Ts.

Insertions generally outnumber deletions in eukaryotic cells[103,113], whereas the opposite is true in *E. coli*[46]. In both cases, large deletions are frequently seen in long repeat tracts[103]. In general, at least two explanations might account for observations of a directional bias in microsatellite mutation. The primary rate of slippage mutation might be higher for insertions than for deletions. Displaced loops might be more easily introduced in the newly synthesized strand (which results in an insertion) than in the template strand. Alternatively, mismatch repair might more easily recognize or more efficiently repair displaced loops on the template strand than on the nascent strand[99]. That mismatch repair is involved in a directional bias is indicated by the fact that mutations in some mismatch-repair genes, such as yeast *MSH3*, differentially affect the rate of insertions and deletions[114]. In *D. melanogaster*, mismatch repair preferentially recognizes and/or corrects primary expansion mutations to leave an excess of contractions and, generally, $(AT)_n$ mutations are repaired more efficiently than $(GT)_n$ changes[115]. If the character of mismatch repair differs between groups of organisms, we might expect consequent differences in microsatellite frequency[92].

TRANSCRIPTION-COUPLED REPAIR
Preferential repair of the transcribed strand of an active gene that is performed by excision-repair pathways.

The future of microsatellites

The evolutionary process of simple repeats is far from simple. One important implication of the complexity of microsatellite evolution is, therefore, that care needs to be taken when using microsatellite data in population genetics studies. For instance, significant mutation-rate heterogeneity among loci means that it might be difficult to translate estimates of genetic distance into absolute timescales. Similarly, directional biases in the mutation process have important consequences for the interpretation of differences in allele size distributions among species, particularly if the character of the bias differs among species[116]. Future mathematical models of microsatellite evolution should therefore aim to incorporate as many of the different forms of mutational heterogeneity as possible. Those who use microsatellites in population genetics studies should select only the markers that are well characterized in terms of mutational properties (mutation rates, directionality, whether all alleles of equal length are identical in sequence), and, preferably, use markers that show uniform rates and patterns. Alternatively, but in many species less realistically, the use of many markers might compensate for heterogeneity in mutational properties among loci.

Microsatellites continue to find their application in areas such as linkage mapping, paternity testing, forensics and for the inference of demographic processes. More recently, they have found most use in linkage-disequilibrium mapping studies, in which associations between markers and trait loci are searched for in population samples[117], and in hitchhiking mapping, in which genome-wide screens for regions that show signs of selection are made[118]. But there are also prospects for new applications. Given their high mutation rate, microsatellites offer a realistic means to study how the overall genomic mutation rate is affected by environmental factors (genetic toxicology). Elevated rates of microsatellite mutations in the germline have been seen in animals and plants that are exposed to ionizing radiation[119,120], and similar observations have been made for minisatellites in humans[121]. Estimating microsatellite mutation rates in samples that are exposed to different forms of radiation or toxic compounds could, when properly set in relation to data from control groups, help to make risk assessments.

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
The first description and analysis of a publicly released assembly of the human genome.
2. Doolittle, W. F. & Sapienza, C. Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601–603 (1980).
3. Orgel, L. E. & Crick, F. H. Selfish DNA: the ultimate parasite. *Nature* **284**, 604–607 (1980).
4. Weber, J. L. Informativeness of human $(dC\text{-}dA)_n$. $(dG\text{-}dT)_n$ polymorphisms. *Genomics* **7**, 524–530 (1990).
5. Weber, J. L. & Wong, C. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**, 1123–1128 (1993).
6. Metzgar, D. & Wills, C. Evidence for the adaptive evolution of mutation rates. *Cell* **101**, 581–584 (2000).
7. Albà, M. M. & Guigó, R. Comparative analysis of amino acid repeats in rodents and humans. *Genome Res.* **14**, 549–545 (2004).
8. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
9. Rat Genome Sequencing Project Consortium. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* **428**, 493–521 (2004).
10. Webster, M. T., Smith, N. G. & Ellegren, H. Microsatellite evolution inferred from human–chimpanzee genomic sequence alignments. *Proc. Natl Acad. Sci. USA* **99**, 8748–8753 (2002).
Provides an unbiased comparison of microsatellite length in humans and chimpanzees on the basis of orthologous genome sequence data.
11. Pascual, M., Schug, M. D. & Aquadro, C. F. High density of long dinucleotide microsatellites in *Drosophila subobscura*. *Mol. Biol. Evol.* **17**, 1259–1267 (2000).
12. Schlotterer, C. & Harr, B. *Drosophila virilis* has long and highly polymorphic microsatellites. *Mol. Biol. Evol.* **17**, 1641–1646 (2000).
13. Hancock, J. M. Simple sequences in a 'minimal' genome. *Nature Genet.* **14**, 14–15 (1996).
14. Toth, G., Gaspari, Z. & Jurka, J. Microsatellites in different eukaryotic genomes: survey and analysis. *Genome Res.* **10**, 967–981 (2000).
15. Katti, M. V., Ranjekar, P. K. & Gupta, V. S. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.* **18**, 1161–1167 (2001).
16. Morgante, M., Hanafey, M. & Powell, W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genet.* **30**, 194–200 (2002).
17. Lagercrantz, U., Ellegren, H. & Andersson, L. The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. *Nucleic Acids Res.* **21**, 1111–1115 (1993).
18. Bachtrog, D., Weiss, S., Zangerl, B., Brem, G. & Schlotterer, C. Distribution of dinucleotide microsatellites in the *Drosophila melanogaster* genome. *Mol. Biol. Evol.* **16**, 602–610 (1999).
19. Arcot, S. S., Wang, Z., Weber, J. L., Deininger, P. L. & Batzer, M. A. Alu repeats: a source for the genesis of primate microsatellites. *Genomics* **29**, 136–144 (1995).
20. Ramsay, L. *et al.* Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J.* **17**, 415–425 (1999).
21. Temnykh, S. *et al.* Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res.* **11**, 1441–1452 (2001).
22. Wilder, J. & Hollocher, H. Mobile elements and the genesis of microsatellites in dipterans. *Mol. Biol. Evol.* **18**, 384–392 (2001).
23. Field, D. & Wills, C. Abundant microsatellite polymorphism in *Saccharomyces cerevisiae*, and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proc. Natl Acad. Sci. USA* **95**, 1647–1652 (1998).
24. Metzgar, D., Thomas, E., Davis, C., Field, D. & Wills, C. The microsatellites of *Escherichia coli*: rapidly evolving repetitive DNAs in a non-pathogenic prokaryote. *Mol. Microbiol.* **39**, 183–190 (2001).
25. Himmelreich, R. *et al.* Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* **24**, 4420–4449 (1996).
26. Ohta, T. & Kimura, M. A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genet. Res.* **22**, 201–204 (1973).
Classic description of the stepwise mutation model.
27. Shriver, M. D., Jin, L., Chakraborty, R. & Boerwinkle, E. VNTR allele frequency distributions under the stepwise mutation model: a computer simulation approach. *Genetics* **134**, 983–993 (1993).
28. Valdes, A. M., Slatkin, M. & Freimer, N. B. Allele frequencies at microsatellite loci: the stepwise mutation model revisited. *Genetics* **133**, 737–749 (1993).
29. Kimmel, M. & Chakraborty, R. Measures of variation at DNA repeat loci under a general stepwise mutation model. *Theor. Popul. Biol.* **50**, 345–367 (1996).
30. Kimmel, M., Chakraborty, R., Stivers, D. N. & Deka, R. Dynamics of repeat polymorphisms under a forward-backward mutation model: within- and between-population variability at microsatellite loci. *Genetics* **143**, 549–555 (1996).
31. Di Rienzo, A. *et al.* Mutational processes of simple-sequence repeat loci in human populations. *Proc. Natl Acad. Sci. USA* **91**, 3166–3170 (1994).
32. Nauta, M. J. & Weissing, F. J. Constraints on allele size at microsatellite loci: implications for genetic differentiation. *Genetics* **143**, 1021–1032 (1996).
33. Feldman, M. W., Bergman, A., Pollock, D. D. & Goldstein, D. B. Microsatellite genetic distances with range constraints: analytic description and problems of estimation. *Genetics* **145**, 207–216 (1997).
34. Pollock, D. D., Bergman, A., Feldman, M. W. & Goldstein, D. B. Microsatellite behavior with range constraints: parameter estimation and improved distances for use in phylogenetic reconstruction. *Theor. Popul. Biol.* **53**, 256–271 (1998).
35. Stefanini, F. M. & Feldman, M. W. Bayesian estimation of range for microsatellite loci. *Genet. Res.* **75**, 167–177 (2000).
36. Garza, J. C., Slatkin, M. & Freimer, N. B. Microsatellite allele frequencies in humans and chimpanzees, with implications for constraints on allele size. *Mol. Biol. Evol.* **12**, 594–603 (1995).
37. Zhivotovsky, L. A. A new genetic distance with application to constrained variation at microsatellite loci. *Mol. Biol. Evol.* **16**, 467–471 (1999).
38. Calabrese, P. & Durrett, R. Dinucleotide repeats in the *Drosophila* and human genomes have complex, length-dependent mutation processes. *Mol. Biol. Evol.* **20**, 715–725 (2003).
39. Cooper, G., Burroughs, N. J., Rand, D. A., Rubinsztein, D. C. & Amos, W. Markov chain Monte Carlo analysis of human Y-chromosome microsatellites provides evidence of biased mutation. *Proc. Natl Acad. Sci. USA* **96**, 11916–11921 (1999).
40. Nielsen, R. & Palsboll, P. J. Single-locus tests of microsatellite evolution: multi-step mutations and constraints on allele size. *Mol. Phylogenet. Evol.* **11**, 477–484 (1999).
41. Renwick, A., Davison, L., Spratt, H., King, J. P. & Kimmel, M. DNA dinucleotide evolution in humans: fitting theory to facts. *Genetics* **159**, 737–747 (2001).
42. Dieringer, D. & Schlotterer, C. Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res.* **13**, 2242–2251 (2003).
43. Bell, G. I. & Jurka, J. The length distribution of perfect dimer repetitive DNA is consistent with its evolution by an unbiased single-step mutation process. *J. Mol. Evol.* **44**, 414–421 (1997).
44. Kruglyak, S., Durrett, R. T., Schug, M. D. & Aquadro, C. F. Equilibrium distributions of microsatellite repeat length resulting from a balance between slippage events and point mutations. *Proc. Natl Acad. Sci. USA* **95**, 10774–10778 (1998).
Provides an integrated model of microsatellite evoluton that takes length mutations as well as base substitutions into account.
45. Calabrese, P. P., Durrett, R. T. & Aquadro, C. F. Dynamics of microsatellite divergence under stepwise mutation and proportional slippage/point mutation models. *Genetics* **159**, 839–852 (2001).

46. Levinson, G. & Gutman, G. A. High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res.* **15**, 5323–5338 (1987).
Firm demonstration of replication slippage as a main mechanism for microsatellite instability.
47. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* **365**, 274–276 (1993).
48. Schlotterer, C. & Tautz, D. Slippage synthesis of simple sequence DNA. *Nucleic Acids Res.* **20**, 211–215 (1992).
Experimental evidence that DNA polymerase is the only enzymatic activity needed for replication slippage.
49. Hile, S. E. & Eckert, K. A. Positive correlation between DNA polymerase α-primase pausing and mutagenesis within polypyrimidine/polypurine microsatellite sequences. *J. Mol. Biol.* **335**, 745–759 (2004).
50. Hauge, X. Y. & Litt, M. A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. *Hum. Mol. Genet.* **2**, 411–415 (1993).
51. Murray, V., Monchawin, C. & England, P. R. The determination of the sequences present in the shadow bands of a dinucleotide repeat PCR. *Nucleic Acids Res.* **21**, 2395–2398 (1993).
52. Shinde, D., Lai, Y., Sun, F. & Arnheim, N. *Taq* DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: $(CA/GT)_n$ and $(A/T)_n$ microsatellites. *Nucleic Acids Res.* **31**, 974–980 (2003).
53. Berg, I., Neumann, R., Cederberg, H., Rannug, U. & Jeffreys, A. J. Two modes of germline instability at human minisatellite *MS1* (locus *D1S7*): complex rearrangements and paradoxical hyperdeletion. *Am. J. Hum. Genet.* **72**, 1436–1447 (2003).
54. Majewski, J. & Ott, J. GT repeats are associated with recombination on human chromosome 22. *Genome Res.* **10**, 1108–1114 (2000).
55. Treco, D. & Arnheim, N. The evolutionarily conserved repetitive sequence $d(TG.AC)_n$ promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis. *Mol. Cell. Biol.* **6**, 3934–3947 (1986).
56. Huang, Q. Y. *et al.* Mutation patterns at dinucleotide microsatellite loci in humans. *Am. J. Hum. Genet.* **70**, 625–634 (2002).
57. Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. & de Knijff, P. Estimating Y chromosome specific microsatellite mutation frequencies using deep rooting pedigrees. *Hum. Mol. Genet.* **6**, 799–803 (1997).
58. Kayser, M. *et al.* Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *Am. J. Hum. Genet.* **66**, 1580–1588 (2000).
59. Hile, S. E., Yan, G. & Eckert, K. A. Somatic mutation rates and specificities at TC/AG and GT/CA microsatellite sequences in nontumorigenic human lymphoblastoid cells. *Cancer Res.* **60**, 1698–1703 (2000).
60. Brohede, J., Primmer, C. R., Moller, A. & Ellegren, H. Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. *Nucleic Acids Res.* **30**, 1997–2003 (2002).
61. Shimoda, N. *et al.* Zebrafish genetic map with 2000 microsatellite markers. *Genomics* **58**, 219–232 (1999).
62. Fitzsimmons, N. N. Single paternity of clutches and sperm storage in the promiscuous green turtle (*Chelonia mydas*). *Mol. Ecol.* **7**, 575–584 (1998).
63. Gardner, M. G., Bull, C. M., Cooper, S. J. & Duffield, G. A. Microsatellite mutations in litters of the Australian lizard *Egernia stokesii*. *J. Evol. Biol.* **13**, 551–560 (2000).
64. Jones, A. G., Rosenqvist, G., Berglund, A. & Avise, J. C. Clustered microsatellite mutations in the pipefish *Syngnathus typhle*. *Genetics* **152**, 1057–1063 (1999).
65. Amos, W., Sawcer, S. J., Feakes, R. W. & Rubinsztein, D. C. Microsatellites show mutational bias and heterozygote instability. *Nature Genet.* **13**, 390–391 (1996).
66. Brinkmann, B., Klintschar, M., Neuhuber, F., Huhne, J. & Rolf, B. Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am. J. Hum. Genet.* **62**, 1408–1415 (1998).
67. Holtkemper, U., Rolf, B., Hohoff, C., Forster, P. & Brinkmann, B. Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. *Hum. Mol. Genet.* **10**, 629–633 (2001).
Introduction of sperm typing as an alternative means for microsatellite mutation detection.
68. Myhre Dupuy, B., Stenersen, M., Egeland, T. & Olaisen, B. Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. *Hum. Mutat.* **23**, 117–124 (2004).
69. Ellegren, H. Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genet.* **24**, 400–402 (2000).
70. Xu, X., Peng, M. & Fang, Z. The direction of microsatellite mutations is dependent upon allele length. *Nature Genet.* **24**, 396–399 (2000).
71. Primmer, C. R., Saino, N., Moller, A. P. & Ellegren, H. Directional evolution in germline microsatellite mutations. *Nature Genet.* **13**, 391–393 (1996).
72. Primmer, C. R., Saino, N., Moller, A. P. and Ellegren, H. Unraveling the process of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. *Mol. Biol. Evol.* **15**, 1047–1054 (1998).
73. Beck, N. R., Double, M. C. & Cockburn, A. Microsatellite evolution at two hypervariable loci revealed by extensive avian pedigrees. *Mol. Biol. Evol.* **20**, 54–61 (2003).
74. Harr, B. & Schlotterer, C. Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short persistence times, which cause their genome-wide underrepresentation. *Genetics* **155**, 1213–1220 (2000).
75. Metzgar, D., Liu, L., Hansen, C., Dybvig, K. & Wills, C. Domain-level differences in microsatellite distribution and content result from different relative rates of insertion and deletion mutations. *Genome Res.* **12**, 408–413 (2002).
76. Estoup, A., Jarne, P. & Cornuet, J. M. Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Mol. Ecol.* **11**, 1591–1604 (2002).
A useful overview of the implications of microsatellite homoplasy in evolutionary studies.
77. Messier, W., Li, S. H. & Stewart, C. B. The birth of microsatellites. *Nature* **381**, 483 (1996).
78. Orti, G., Pearse, D. E. & Avise, J. C. Phylogenetic assessment of length variation at a microsatellite locus. *Proc. Natl Acad. Sci. USA* **94**, 10745–10749 (1997).
79. Angers, B. & Bernatchez, L. Complex evolution of a salmonid microsatellite locus and its consequences in inferring allelic divergence from size information. *Mol. Biol. Evol.* **14**, 230–238 (1997).
80. Primmer, C. R. & Ellegren, H. Patterns of molecular evolution in avian microsatellites. *Mol. Biol. Evol.* **15**, 997–1008 (1998).
81. Harr, B., Zangerl, B. & Schlotterer, C. Removal of microsatellite interruptions by DNA replication slippage: phylogenetic evidence from *Drosophila*. *Mol. Biol. Evol.* **17**, 1001–1009 (2000).
82. Zhu, Y., Strassmann, J. E. & Queller, D. C. Insertions, substitutions, and the origin of microsatellites. *Genet. Res.* **76**, 227–236 (2000).
83. Taylor, J. S., Durkin, J. M. & Breden, F. The death of a microsatellite: a phylogenetic perspective on microsatellite interruptions. *Mol. Biol. Evol.* **16**, 567–572 (1999).
84. Schlotterer, C., Amos, B. & Tautz, D. Conservation of polymorphic simple sequence loci in cetacean species. *Nature* **354**, 63–65 (1991).
85. Colson, I. & Goldstein, D. B. Evidence for complex mutations at microsatellite loci in *Drosophila*. *Genetics* **152**, 617–627 (1999).
86. Ellegren, H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet.* **16**, 551–558 (2000).
87. Schug, M. D., Mackay, T. F. & Aquadro, C. F. Low mutation rates of microsatellite loci in *Drosophila melanogaster*. *Nature Genet.* **15**, 99–102 (1997).
Characterizaton of mutation rate and repeat lengths of microsatellites in *D. melanogaster*, revealing significant differences from vertebrate genomes.
88. Leopoldino, A. M. & Pena, S. D. The mutational spectrum of human autosomal tetranucleotide microsatellites. *Hum. Mutat.* **21**, 71–79 (2003).
89. Sibly, R. M. *et al.* The structure of interrupted human AC microsatellites. *Mol. Biol. Evol.* **20**, 453–459 (2003).
90. Crozier, R. H., Kaufmann, B., Carew, M. E. & Crozier, Y. C. Mutability of microsatellites developed for the ant *Camponotus consobrinus*. *Mol. Ecol.* **8**, 271–276 (1999).
91. Glenn, T. C., Stephan, W., Dessauer, H. C. & Braun, M. J. Allelic diversity in alligator microsatellite loci is negatively correlated with GC content of flanking sequences and evolutionary conservation of PCR amplifiability. *Mol. Biol. Evol.* **13**, 1151–1154 (1996).
92. Bachtrog, D., Agis, M., Imhof, M. & Schlotterer, C. Microsatellite variability differs between dinucleotide repeat motifs-evidence from *Drosophila melanogaster*. *Mol. Biol. Evol.* **17**, 1277–1285 (2000).
93. Harr, B., Zangerl, B., Brem, G. & Schlotterer, C. Conservation of locus-specific microsatellite variability across species: a comparison of two *Drosophila* sibling species, *D. melanogaster* and *D. simulans*. *Mol. Biol. Evol.* **15**, 176–184 (1998).
94. Mellon, I., Rajpal, D. K., Koi, M., Boland, C. R. & Champe, G. N. Transcription-coupled repair deficiency and mutations in human mismatch repair genes. *Science* **272**, 557–560 (1996).
95. Ellegren, H., Smith, N. G. & Webster, M. T. Mutation rate variation in the mammalian genome. *Curr. Opin. Genet. Dev.* **13**, 562–568 (2003).
96. Santibanez-Koref, M. F., Gangeswaran, R. & Hancock, J. M. A relationship between lengths of microsatellites and nearby substitution rates in mammalian genomes. *Mol. Biol. Evol.* **18**, 2119–2123 (2001).
Offers a suggestion for how variation in microsatellite length might relate to point-mutation-rate heterogeneity.
97. Brohede, J., Arnheim, N. & Ellegren, H. Single molecule analysis of the hypermutable tetranucleotide repeat locus D21S1245 through sperm genotyping: a heterogeneous pattern of mutation but no clear male age effect. *Mol. Biol. Evol.* **21**, 58–64 (2004).
98. Henderson, S. T. & Petes, T. D. Instability of simple sequence DNA in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**, 2749–2757 (1992).
Introduces an experimental approach to the study of microsatellite mutations, using artificial plasmid-borne repeat sequences associated with a resistance or reporter gene.
99. Sia, E. A., Kokoska, R. J., Dominska, M., Greenwell, P. & Petes, T. D. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. *Mol. Cell. Biol.* **17**, 2851–2858 (1997).
100. Strauss, B. S., Sagher, D. & Acharya, S. Role of proofreading and mismatch repair in maintaining the stability of nucleotide repeats in DNA. *Nucleic Acids Res.* **25**, 806–813 (1997).
101. Tran, H. T., Keen, J. D., Kricker, M., Resnick, M. A. & Gordenin, D. A. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol. Cell. Biol.* **17**, 2859–2865 (1997).
102. Gutierrez, P. J. & Wang, T. S. Genomic instability induced by mutations in *Saccharomyces cerevisiae POL1*. *Genetics* **165**, 65–81 (2003).
103. Wierdl, M., Dominska, M. & Petes, T. D. Microsatellite instability in yeast: dependence on the length of the microsatellite. *Genetics* **146**, 769–779 (1997).
104. Yamada, N. A. *et al.* Relative rates of insertion and deletion mutations in dinucleotide repeats of various lengths in mismatch repair proficient mouse and mismatch repair deficient human cells. *Mutat. Res.* **499**, 213–225 (2002).
105. Petes, T. D., Greenwell, P. W. & Dominska, M. Stabilization of microsatellite sequences by variant repeats in the yeast *Saccharomyces cerevisiae*. *Genetics* **146**, 491–498 (1997).
106. Bacon, A. L., Farrington, S. M. & Dunlop, M. G. Sequence interruptions confer differential stability at microsatellite alleles in mismatch repair-deficient cells. *Hum. Mol. Genet.* **9**, 2707–2713 (2000).
107. Maurer, D. J., O'Callaghan, B. L. & Livingston, D. M. Orientation dependence of trinucleotide CAG repeat instability in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **16**, 6617–6622 (1996).
108. Eckert, K. A. & Yan, G. Mutational analyses of dinucleotide and tetranucleotide microsatellites in *Escherichia coli*: influence of sequence on expansion mutagenesis. *Nucleic Acids Res.* **28**, 2831–2838 (2000).
109. Lee, J. S., Hanford, M. G., Genova, J. L. & Farber, R. A. Relative stabilities of dinucleotide and tetranucleotide repeats in cultured mammalian cells. *Hum. Mol. Genet.* **8**, 2567–2572 (1999).
110. Sagher, D., Hsu, A. & Strauss, B. Stabilization of the intermediate in frameshift mutation. *Mutat. Res.* **423**, 73–77 (1999).
111. Boyer, J. C. *et al.* Sequence dependent instability of mononucleotide microsatellites in cultured mismatch repair proficient and deficient mammalian cells. *Hum. Mol. Genet.* **11**, 707–713 (2002).
112. Harfe, B. D. & Jinks-Robertson, S. Sequence composition and context effects on the generation and repair of frameshift intermediates in mononucleotide runs in *Saccharomyces cerevisiae*. *Genetics* **156**, 571–578 (2000).
113. Twerdi, C. D., Boyer, J. C. & Farber, R. A. Relative rates of insertion and deletion mutations in a microsatellite sequence in cultured cells. *Proc. Natl Acad. Sci. USA* **96**, 2875–2879 (1999).
114. Strand, M., Earley, M. C., Crouse, G. F. & Petes, T. D. Mutations in the *MSH3* gene preferentially lead to deletions within tracts of simple repetitive DNA in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **92**, 10418–10421 (1995).
115. Harr, B., Todorova, J. & Schlotterer, C. Mismatch repair-driven mutational bias in *D. melanogaster*. *Mol. Cell* **10**, 199–205 (2002).
116. Amos, W., Hutter, C. M., Schug, M. D. & Aquadro, C. F. Directional evolution of size coupled with ascertainment bias for variation in *Drosophila* microsatellites. *Mol. Biol. Evol.* **20**, 660–662 (2003).
117. Ohashi, J. & Tokunaga, K. Power of genome-wide linkage disequilibrium testing by using microsatellite markers. *J. Hum. Genet.* **48**, 487–491 (2003).

118. Schlotterer, C. Hitchhiking mapping — functional genomics from the population genetics perspective. *Trends Genet.* **19**, 32–38 (2003).
119. Ellegren, H., Lindgren, G., Primmer, C. R. & Moller, A. P. Fitness loss and germline mutations in barn swallows breeding in Chernobyl. *Nature* **389**, 593–596 (1997).
120. Kovalchuk, O., Kovalchuk, I., Arkhipov, A., Hohn, B. & Dubrova, Y. E. Extremely complex pattern of microsatellite mutation in the germline of wheat exposed to the post-Chernobyl radioactive contamination. *Mutat. Res.* **525**, 93–101 (2003).
121. Dubrova, Y. E. *et al.* Human minisatellite mutation rate after the Chernobyl accident. *Nature* **380**, 683–686 (1996).
122. Spritz, R. A. Duplication/deletion polymorphism 5′- to the human β-globin gene. *Nucleic Acids Res.* **9**, 5037–5047 (1981).
123. Miesfeld, R., Krystal, M. & Arnheim, N. A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ- and β-globin genes. *Nucleic Acids Res.* **9**, 5931–5947 (1981).
124. Hamada, H. & Kakunaga, T. Potential Z-DNA forming sequences are highly dispersed in the human genome. *Nature* **298**, 396–398 (1982).
125. Jeffreys, A. J., Wilson, V. & Thein, S. L. Hypervariable 'minisatellite' regions in human DNA. *Nature* **314**, 67–73 (1985).
126. Tautz, D., Trick, M. & Dover, G. A. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* **322**, 652–656 (1986).
127. Litt, M. & Luty, J. A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. *Am. J. Hum. Genet.* **44**, 397–401 (1989).
This paper, and references 128 and 129, introduce the use of PCR for genotyping microsatellites.
128. Weber, J. L. & May, P. E. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *Am. J. Hum. Genet.* **44**, 388–396 (1989).
129. Tautz, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. *Nucleic Acids Res.* **17**, 6463–6471 (1989).
130. Ellegren, H. DNA typing of museum birds. *Nature* **354**, 113 (1991).
131. Weissenbach, J. *et al.* A second-generation linkage map of the human genome. *Nature* **359**, 794–801 (1992).
132. Bowcock, A. M. *et al.* High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**, 455–457 (1994).
133. Dawid, A. P., Mortera, J. & Pascali, V. L. Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing. *Forensic Sci. Int.* **124**, 55–61 (2001).
134. Whittaker, J. C. *et al.* Likelihood-based estimation of microsatellite mutation rates. *Genetics* **164**, 781–787 (2003).

Acknowledgements

The author would like to acknowledge two particularly helpful reviewers who provided useful comments on the manuscript. The author's work was supported in part by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning.

Competing interests statement

The author declares that he has no competing financial interests.

Online links

DATABASES

The following terms in this article are linked online to:
Entrez: http://www.ncbi.nih.gov/Entrez
MSH3

FURTHER INFORMATION

RepeatMasker: http://www.repeatmasker.org
Sputnik: http://espressosoftware.com/pages/sputnik.jsp
Tandem Repeats Finder: http://c3.biomath.mssm.edu/trf.html
Access to this links box is available online.