Review

Measurements of spontaneous rates of mutations in the recent past and the near future

Fyodor A. Kondrashov^{1,*} and Alexey S. Kondrashov²

¹*Bioinformatics and Genomics Programme, Centre for Genomic Regulation, C/Dr. Aiguader 88, Barcelona Biomedical Research Park Building 08003, Barcelona, Spain*

²*Evolutionary Biology and Ecology Department, Life Sciences Institute, University of Michigan, Ann Arbor, MI 48109, USA*

The rate of spontaneous mutation in natural populations is a fundamental parameter for many evolutionary phenomena. Because the rate of mutation is generally low, most of what is currently known about mutation has been obtained through indirect, complex and imprecise methodological approaches. However, in the past few years genome-wide sequencing of closely related individuals has made it possible to estimate the rates of mutation directly at the level of the DNA, avoiding most of the problems associated with using indirect methods. Here, we review the methods used in the past with an emphasis on next generation sequencing, which may soon make the accurate measurement of spontaneous mutation rates a matter of routine.

Keywords: mutation; sequencing; estimating mutation rate; mutation accumulation

1. INTRODUCTION

Mutation, as the ultimate source of all genetic variations, is arguably among the most fundamental processes in biology, and the mutation rate is one of the key quantities that describe life. Estimating the rates of different kinds of mutations and characterizing the mechanisms of their origin are important ingredients for understanding many other biological phenomena, such as the evolution of sex or genetic disease. However, aside from these practical implications, an accurate knowledge of the mutation rate has an inherent value in expanding our fundamental understanding of the natural world.

Mutations have been the subject of quantitative research for quite some time (Danforth 1923; Muller 1927; Haldane 1935; Timofeeff-Ressovsky 1935; Mukai 1964; Mukai *et al.* 1972; Crow & Abrahamson 1997) and there have been many attempts at estimating the rates of spontaneous mutations (for a review, see Drake 1993; Drake *et al.* 1998; Lynch *et al.* 1999, 2006; Lynch 2006; Baer *et al.* 2007). The efforts of the past century have produced a fragmented picture. On the one hand, our level of awareness of the importance of mutations is phenomenal, embracing various branches of biology from the mechanisms of mutation to its impacts on fitness and disease. On the other hand, study of spontaneous mutation has barely reached a level of accuracy where it can be

considered a quantitative science. Why are we still struggling to make accurate measurements of this fundamental parameter, 100 years after the mass of the electron and the speed of light were determined?

There are six reasons why measuring the rate of mutation has proven to be so difficult. First and foremost, it is because the per-nucleotide rate of mutation is extremely low, especially in eukaryotes (Drake *et al.* 1998; Lynch 2006; Lynch *et al.* 2006), but can be high in viruses (Drake *et al.* 1998; Gago *et al.* 2009). To measure such low rates, extremely careful and laborious methods are required. Second, mutation is a random process and there is no way to know where and when a mutation will occur. Third, there are several kinds of mutations, including point mutations (single-nucleotide substitutions, insertions, deletions, and complex events) and large-scale mutations (long insertions and deletions, duplications and polyploidizations, inversions, results of gene conversion, transpositions and long complex events; Petrov *et al.* 2000; Yang *et al.* 2001a,b; Ball *et al.* 2005; Chen *et al.* 2005; Hastings *et al.* 2009) and their rates need to be measured separately. Fourth, the rate of mutation can depend on the genetic background, especially on the context surrounding the site in question or its methylation (Holliday & Grigg 1993; Rogozin & Pavlov 2003). Thus, simply determining the average rate even of a particular kind of mutation is not enough: ideally an accurate measurement would also take into account the frequency of different contexts in the genome. Fifth, the mutation rate can be influenced dramatically by many mutagenic environmental and physiological factors, including temperature (Muller 1932), chemicals

* Author for correspondence (fyodor.kondrashov@crg.es).

One contribution of 16 to a Theme Issue 'The population genetics of mutations: good, bad and indifferent' dedicated to Brian Charlesworth on his 65th birthday.

(compilation of hazardous chemicals (HAZARD DB 2009) lists approx. 700 known mutagens), radiation, age, sex, etc. Finally, the mutation rate itself is subject to selection and evolution, with some authors suggesting that some organisms increase the rate of mutation in times of stress (Sniegowski *et al.* 1997; Loewe *et al.* 2003; Avila *et al.* 2006; Lynch 2008).

2. METHODS FOR DETERMINING MUTATION RATES

Several authors have classified conceptual approaches of measuring spontaneous mutation rates (Kondrashov 1998; Fu & Huai 2003; Baer *et al.* 2007) while others have collected qualitative estimates of mutation rates (Drake *et al.* 1998; Keightley & Eyre-Walker 1999; Lynch 2006; Lynch *et al.* 2006; Keightley & Halligan 2009). However, these approaches and estimates, in our opinion, are rapidly becoming obsolete, at least in the context of measuring spontaneous mutation rates at the sequence level. Therefore, we present only a brief overview of these estimates, not for the purpose of guidance in selecting one of these approaches for further usage, but as a tribute to the enormous efforts that generations of scientists have invested in studying spontaneous mutation.

(a) *Phenotypic mutation screening*

A straightforward approach to estimate the rate of mutation is to screen for phenotypically evident mutants (for a review, see Drake *et al.* 1998; Fu & Huai 2003). Then, to estimate the per nucleotide rate of mutation it is necessary to know the mutational target, the length of the sequence which, when disrupted, leads exclusively to the screened phenotype. The application of this method has been hindered in the past by a poor knowledge of the fraction of functional sites per locus (Kondrashov 1998), but the availability of complete genomes has eliminated this problem. This approach has been applied to humans (Sommer 1995; Sommer & Ketterling 1996; Kondrashov 2003; Nachman 2004), *Drosophila* (Drake *et al.* 1998) and other species (Schrage *et al.* 1999; Mackwan *et al.* 2008), but has never been very popular for eukaryotes, with only one large-scale study performed in recent years (Yang *et al.* 2001a). In general, phenotype screening almost certainly underestimates mutation rates, either because not all nonsense mutations produce a phenotype (Nachman 2004) and available corrections (Drake *et al.* 1998) are not perfect, or because of the difficulties arising from clusters of mutants (Yang *et al.* 2001a; Fu & Huai 2003; Gong *et al.* 2005a,b), or because a mutant can be easily overlooked in the course of screening. The main problem, however, with this type of approach is the sheer volume of work involved in phenotype screens owing to the low spontaneous mutation rates. For example, Yang *et al.* (2001a) screened approximately 900 000 flies and found only 16 mutation events at eight loci.

(b) *Measuring the fitness impact of accumulating mutations*

Measuring the fitness effect of mutations accumulating over time in a laboratory population under relaxed or

absent selection has been a popular choice for estimating the rate of deleterious mutation (Mukai 1964; Mukai *et al.* 1972; Charlesworth *et al.* 1990, 1994; Kondrashov & Houle 1994; Fernández & López-Fanjul 1996; Kibota & Lynch 1996; Keightley & Caballero 1997; Schultz *et al.* 1999; Vassilieva & Lynch 1999; Loewe *et al.* 2003; Charlesworth *et al.* 2004; Houle & Nuzhdin 2004; Gong *et al.* 2005a,b; Roles & Conner 2008). Starting from an inbred population, it is possible to measure the impact on fitness of many mutations that have accumulated over dozens of generations. Thus, this approach measures the rate of deleterious mutations and, if the deleterious fraction of the total mutation rate can be estimated, then one can deduce the total rate. This approach, however, has a variety of problems and its application remains controversial (Keightley & Eyre-Walker 1999; Fry 2001). An inbred population may not reflect mutation patterns expected in a wild population. Also, a high rate of homozygosity can largely obscure selection and makes it difficult to maintain the population in the laboratory, although the latter issue is not as important in selfing organisms. One way to avoid these problems is to maintain a small, randomly mating population that is not inbred, the so-called 'middle class neighbourhood' population (Shabalina *et al.* 1997). Nevertheless, other problems are unavoidable, including the principal impossibility of eliminating selection completely even under laboratory conditions, difficulties in detecting very mildly deleterious mutations, and the possibility of mutations affecting the rate of mutation itself in the course of the many generations under relaxed selection. Therefore, it is not surprising that different authors obtained very different estimates of the mutation rate using this approach. Of course, this approach is also very time-consuming, requiring constant vigilance to maintain the inbred population, to prevent contamination and to repeatedly assay fitness.

(c) *Measuring the rate of evolution at neutral sites*

The rate of neutral evolution is equal to the mutation rate (Kimura 1968). This simple relationship has been used several times to estimate the spontaneous rate of mutation (Crow 1993; Drake *et al.* 1998; Nachman & Crowell 2000; Kumar & Subramanian 2002; Silva & Kondrashov 2002; Mitchell & Graur 2005), with varying degrees of success. In particular, it has been suggested that contrasting the rate of evolution of neutral sequences with that of sequences under selection can be used to estimate the rate of deleterious mutation (Kondrashov & Crow 1993). Several such estimates have since appeared (Eyre-Walker & Keightley 1999; Keightley & Eyre-Walker 2000; Cutter & Payseur 2003), yet the accuracy of the underlying rate of spontaneous mutations in these estimates has been questioned, most notably on account of the usage of synonymous sites as neutral markers (Kondrashov 2001). Two problems are apparent with this method. Firstly, as exemplified by the use of synonymous sites, there must be a good reason to assume neutrality of particular sequences. Secondly, the rate of neutral evolution should not be measured

by comparing distant species owing to the difficulty of estimating the number of hidden multiple substitutions per site (Li 1997) and a lack of knowledge of the true number of generations separating the two species. The study of Nachman & Crowell (2000) avoided both of these issues by selecting orthologous pseudogenes of humans and chimps and, as a result, their estimate remains the best application of this approach.

(d) Inferring parameters of mutation from data on within-population variation

Many properties of equilibrium populations depend on the mutation rates and, thus, can be used to estimate them. In particular, nucleotide diversity at the sequence level is $4N_e\mu$ in diploids, and thus can be used to estimate μ (Deng & Lynch 1996; Kondrashov 1998; Li & Deng 2005; Deng *et al.* 2006; Messer 2009). However, because N_e is not known independently, this relationship is more useful for estimating N_e and not μ (Kondrashov 1998; Caballero 2006). A more promising approach is to estimate the rate of deleterious mutation from the data on inbreeding depression in highly selfed populations, as this is mainly owing to recent mutations and thus does not depend on the intensity of genetic drift (Charlesworth *et al.* 1990).

(e) Direct sequence-level screening of mutations accumulated in the course of 10–1000 generations

In recent years, several studies have been published in which the rate of spontaneous mutation was measured directly at the DNA level (Denver *et al.* 2000, 2004; Malpica *et al.* 2002; Howell *et al.* 2003; Haag-Liautard *et al.* 2007, 2008; Lynch *et al.* 2008; Keightley *et al.* 2009). Three of them (Lynch *et al.* 2008; Keightley *et al.* 2009; Xue *et al.* 2009) used a direct sequencing approach, demonstrating that direct sequencing is now accurate enough to be used to estimate even low rates of mutations (Kondrashov 2008). Most of these studies have sequenced a mutation accumulation line, with the exception of those of Howell *et al.* (2003) and Xue *et al.* (2009), who sequenced a section of the mitochondrial genome and Y chromosome, respectively, traced through known pedigrees of approximately 10 generations. Pedigree data essentially replace mutation accumulation lines in the human population, and although the use of non-recombining loci simplifies their use by eliminating the possibility of contamination from unrelated individuals, data obtained for the human mitochondria (Howell *et al.* 2003) remain controversial (Henn *et al.* 2009).

(f) Applying mutagens

One of the easiest approaches is to apply artificial mutagens. If the ratio of the spontaneous to mutagen-induced mutation rates is known *a priori*, then measuring the mutagen-induced rate of mutation can be used to estimate the rate of spontaneous mutation. The mutagenic treatment can greatly increase the rate of mutation, thereby reducing the amount of work. While artificial mutagenesis has long been used to

obtain mutants and estimate the impact of mutation on quantitative traits (Mukai 1970; Ohnishi 1977), only two studies on estimating the spontaneous rate of mutations have been performed (Keightley & Ohnishi 1998; Yang *et al.* 2001b). A potentially resolvable issue with this approach is the difficulty of quantifying the impact of the mutagen. A more fatal issue is that, given the nature of the mutagen, it is likely to have profoundly different effects on different kinds of mutations, such as point mutations and indels, and, therefore, cannot be used to obtain good estimates of all spontaneous mutations.

3. A SNAPSHOT OF CURRENT KNOWLEDGE

What do we know about spontaneous rates of mutation? It seems that the problems with measuring rate of mutations have up until very recently thwarted direct quantitative measurements. Measurements of more qualitative factors of mutation have been much more successful. We know that transitions are usually (Keller *et al.* 2007) more frequent than transversions (Kondrashov 2003; Denver *et al.* 2004; Haag-Liautard *et al.* 2007; Lynch *et al.* 2008). The rate of mutation is dependent on the number of germline cell divisions per generation (Drost & Lee 1995; Drake *et al.* 1998), which implies an accumulation of mutations in older individuals and more mutations in males owing to more pre-meiotic cell divisions (Crow 2000; Makova & Li 2002; but see Crow 2006) and that intra-spermatogonial selection can alter the rate of germline mutation (Crow 2006; Arnheim & Calabrese 2009). We have a good understanding of the mechanisms involved in spontaneous mutations, including recombination (Hastings *et al.* 2009), various repair mechanisms (Wyman & Kanaar 2006; Lewin 2008, ch. 20) and of context dependence, the influence of the nucleotide state of neighbouring sites (Gordenin & Resnick 1998; Kondrashov & Rogozin 2004; Kondrashov *et al.* 2006; Hodgkinson *et al.* 2009). Different polymerases can lead to different patterns of mutation (Pavlov *et al.* 2006; Arana *et al.* 2008) and their function also causes the mutagenic effects of methylation (Holliday & Grigg 1993) and mutational hotspots (Rogozin & Pavlov 2003). We also have some estimates of the relative rates of indels (Petrov *et al.* 2000; Yang *et al.* 2001a,b; Denver *et al.* 2004; Haag-Liautard *et al.* 2007), transposon insertions (Nuzhdin & Mackay 1994) and strand biases in the mitochondrial (Tanaka & Ozawa 1994), nuclear and prokaryotic genomes (Green *et al.* 2003; Morton & Morton 2007; Unniraman & Schatz 2007). Finally, the deleterious mutation rate is likely to be $U > 1$ for many eukaryotic species with measured mutation rates (Kondrashov 2001, 2003; Denver *et al.* 2004; Haag-Liautard *et al.* 2007).

As far as quantitative estimates of mutation rate are concerned, a large fraction of them have been obtained in recent studies that use a multigeneration direct sequencing approach (table 1). Other methods have also produced quantitative estimates (for a review, see Drake *et al.* 1998; Lynch 2006; Lynch *et al.* 2006), but, owing to the inherent caveats associated with all approaches other than direct DNA sequencing, these estimates are uncertain and must be repeated. A

Table 1. Direct sequencing estimates of mutation rates. Two probably accurate but indirect estimates (Nachman & Crowell 2000; Kondrashov 2003) are included for *Homo sapiens*.

species	rate of mutation per site per generation	comments	references
<i>Homo sapiens</i>	2.5×10^{-8} total (from species divergence) 1.8×10^{-8} total (phenotypic mutations analysis) 3.0×10^{-8} total (direct sequencing of Y chromosome)	indels are rare	Nachman & Crowell (2000), Kondrashov (2003), Xue <i>et al.</i> (2009)
<i>Drosophila melanogaster</i>	8.4×10^{-9} total 3.5×10^{-9} to 5.8×10^{-9} point mutations	some variation of rates between different lines; supports higher rate estimates	Haag-Liautard <i>et al.</i> (2007), Keightley <i>et al.</i> (2009)
<i>Drosophila melanogaster</i> mitochondrion	7.2×10^{-8} total 6.2×10^{-8} point mutations	high rate of G→A mutations on the major strand; higher than estimates from species divergence	Haag-Liautard <i>et al.</i> (2008)
<i>Caenorhabditis elegans</i>	2.1×10^{-8} total 9.1×10^{-9} point mutations	insertions very common; higher than previous estimates	Denver <i>et al.</i> (2004)
<i>Caenorhabditis elegans</i> mitochondrion	1.6×10^{-7} total 9.7×10^{-8} point mutations	higher than previous estimates	Denver <i>et al.</i> (2000)
<i>Saccharomyces cerevisiae</i>	0.33×10^{-9}	similar to previous estimates	Lynch <i>et al.</i> (2008)
<i>S. cerevisiae</i> mitochondrion	1.2×10^{-8} point mutations (average of two methods) 7.5×10^{-9} indels (average of two methods)		Lynch <i>et al.</i> (2008)

symptom of the uncertainty of indirect estimates is the large variability of estimates obtained by different authors, especially for *Drosophila melanogaster*. The direct DNA sequencing methods have yet to be independently replicated, and perhaps the estimates reported in table 1 should also be treated with caution. However, direct sequencing appears to be associated with low error rates; in particular Lynch *et al.* (2008) and Keightley *et al.* (2009) have not found a single false-positive mutation, indicating that the false positive rate of direct sequencing is lower than the limit of accuracy. In addition, it is heartening to see that the studies that have obtained these estimates have placed the accurate measurement of the spontaneous mutation rate, not its implications, as the primary goal.

4. PERSPECTIVES

(a) *The dawn of a new era in measuring spontaneous mutation rates*

The use of mutation accumulation over the course of many generations, which may lead to an underestimation of the rate of mutation owing to the action of selection in the laboratory (Kondrashov & Houle 1994), is the last bastion that is yet to fall before the advance of sequencing technology. The 1000-dollar human genome (Mardis 2006), if realized, will make sequencing cheap and accurate enough to eliminate

the need to accumulate mutation for many generations, which is obviously impossible for any organism with a not-too-short generation time. Even if the cost of genome sequencing is currently not as cheap as advertised for the future it is doubtful that at present new mutation accumulation lines will be started for estimating spontaneous mutation rates; by the time enough generations pass, genome sequencing may already be cheap enough to sequence individuals separated by very few generations.

In the near future, researchers may sequence randomly selected individuals from parental (P), and two generations of the offspring generations, F₁ and F₂. The sequencing of individuals from the F₂ is required to eliminate somatic mutations, and several such triplet generations could be sequenced to obtain accurate measurements. If the per nucleotide mutation rate is approximately 10^{-8} , to obtain approximately 100 mutations it is necessary to sequence approximately 10^{11} nucleotides (estimated from two P genotypes, one F₁ and seven F₂ genotypes for one generation triplet). A key fact is the apparent accuracy of even the current versions of the next generation sequencing methods, with Lynch *et al.* (2008) and Keightley *et al.* (2009) reporting no false positives among approximately 5×10^6 nucleotides sequenced with $3\times$ or greater coverage and approximately 7.2×10^7 nucleotides sequenced with $5\times$ or greater coverage, respectively.

(b) *Unanswered questions concerned with spontaneous mutation rates*

As powerful as direct sequencing may be, it has not yet reached a level of economy to be able to answer all fundamental questions pertinent to mutation rates. The four recent studies that involved direct sequencing have yielded very few mutations: four in the human pedigree (Xue *et al.* 2009), 35 in *Saccharomyces cerevisiae* (Lynch *et al.* 2008), 37 (Haag-Liautard *et al.* 2007) and 174 (Keightley *et al.* 2009) in *D. melanogaster* and 30 in *Caenorhabditis elegans* (Denver *et al.* 2004). Although some qualitative observations on different kinds of mutations have been made in some of these studies (table 1), obviously, to be of use for refining this knowledge on a quantitative level, the amount of data must increase by more than an order of magnitude. Nevertheless, it may be useful to define a set of unanswered fundamental questions pertaining to the study of mutation rates that may yet be tackled with the advance of cheaper sequencing.

At the heart of all questions, secondary to the average genomic mutation rate is the issue of its natural variability. The first set of questions deals with the intraspecies and interspecies variation of mutation rates. How different are the rates of mutations between species? To some degree a crack at this question has been attempted with the study of selfing versus obligatory sexual organisms (Charlesworth *et al.* 1994; Kondrashov 2001; Denver *et al.* 2004), but it must be set broader than just a question of life cycle. A study of the mutation rate of closely related species will give an idea of how quickly the rate of mutation evolves. Similarly, the distribution of the rates of mutation within a population is intriguing and if such mutational variability is high, then the second type of questions can be of fundamental importance: what is the impact of environmental factors on the rate of mutation? Is the discharge of anthropogenic mutagens into the environment having an effect on the mutation rate of various species? Is there a difference in the mutation rate between populations of the same species living in different environments, for example, hot versus cold? Finally, the question of intragenomic mutation rate variability must be addressed. Methylation and transition/transversion bias surely have a large impact on such variability, but the more detailed impact of context, chromosomal location or other uncharacterized factors is uncertain (Kondrashov *et al.* 2006; Hodgkinson *et al.* 2009). Similarly, precise measurements of the rate of different kinds of mutations, such as indels and substitutions, need to be performed.

(c) *Impact of mutation at the phenotype and fitness levels*

While new-generation sequencing will produce good quantitative description of the mutation process at the sequence level in the not-too-distant future, this will not lead automatically to understanding of the impact of mutation at the higher level of organization. What fractions of de novo mutations are deleterious, neutral and advantageous (Keightley & Eyre-Walker 2010; Orr 2010; Trindade *et al.* 2010)? How rapidly

do fresh mutations supply genetic variation of quantitative traits, i.e. how large is the mutational heritability of various traits (Lynch *et al.* 1999)? What is the magnitude of mutational covariation of traits and fitness? Even a complete knowledge of mutation at the sequence level will not be enough to address these questions, because the answers to them depend on the nature of genotype > phenotype (including genotype > fitness) map. Thus, accumulation of mutations in the course of many generations may still be used to study the phenotypic impacts of mutations. Still, indirect estimates of the genomic deleterious mutation rate based on the fraction of non-neutral sites in the genome may be more promising, because selection at many sites is too weak to be detected in the course of a mutation accumulation study.

(d) *Conclusion*

In the recent past, two circumstances have inhibited a more in-depth study of spontaneous rates of mutation. First, it was the natural difficulties that arise in measuring such infrequent random events. Even a decade ago, direct sequencing of entire genomes of closely related individuals to measure the mutation rate was inconceivable owing to the high cost and low accuracy, although acknowledged as the most problem-free approach of all possibilities (Kondrashov 1998). Two such attempts have already been made (Lynch *et al.* 2008; Keightley *et al.* 2009) and we have every reason to believe that many more estimates obtained in a similar manner will follow.

The second hindering circumstance is that many scientists were more focused on the implications of spontaneous mutations, such as the evolution of sex, and for this purpose have been estimating the rate of mutation indirectly (Drake *et al.* 1998; Kondrashov 2008). At the same time, few studies were devoted exclusively to improving the accuracy and cost effectiveness of such measurements. Improvement in the cost and accuracy of next generation sequencing technologies in the near future may allow the accurate estimation of spontaneous mutation rates in many species through direct sequencing of closely related individuals. Such estimates can be accelerated by a conscious community-driven effort, similar in spirit to that of the Human Genome Project or the CERN Large Hadron Collider, aimed at the scrupulous measurements of one of the most fundamental parameters in evolution—the rate of spontaneous mutation.

This paper is dedicated on the occasion of the 65th birthday of Brian Charlesworth who made a number of contributions to measurements of spontaneous mutation rates, both through the impact of mutation on fitness and by direct sequence level-data, as well as to theoretical investigation of the impact of spontaneous mutation on the structure of the genome and genetic systems.

REFERENCES

- Arana, M. E., Seki, M., Wood, R. D., Rogozin, I. B. & Kunkel, T. A. 2008 Low-fidelity DNA synthesis by human DNA polymerase θ . *Nucl. Acid. Res.* **36**, 3847–3856. (doi:10.1093/nar/gkn310)

- Arnheim, N. & Calabrese, P. 2009 Understanding what determines the frequency and pattern of human germline mutations. *Nat. Rev. Genet.* **10**, 478–488. (doi:10.1038/nrg2529)
- Avila, V., Chavarrias, D., Sánchez, E., Manrique, A., López-Fanjul, C. & García-Dorado, A. 2006 Increase of the spontaneous mutation rate in a long-term experiment with *Drosophila melanogaster*. *Genetics* **173**, 267–277. (doi:10.1534/genetics.106.056200)
- Baer, C. F., Miyamoto, M. M. & Denver, D. R. 2007 Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat. Rev. Genet.* **8**, 619–631. (doi:10.1038/nrg2158)
- Ball, E. V., Stenson, P. D., Abeyasinghe, S. S., Krawczak, M., Cooper, D. N. & Chuzhanova, N. A. 2005 Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity. *Hum. Mutat.* **26**, 205–213. (doi:10.1002/humu.20212)
- Caballero, A. 2006 Estimation of the upper limit of the mutation rate and mean heterozygous effect of deleterious mutations. *Genet. Res.* **88**, 137–141. (doi:10.1017/S001667230700849X)
- Charlesworth, B., Charlesworth, D. & Morgan, M. T. 1990 Genetic loads and estimates of mutation rates in highly inbred plant populations. *Nature* **347**, 380–382. (doi:10.1038/347380a0)
- Charlesworth, D. E., Lyons, E. & Litchfield, L. B. 1994 Inbreeding depression in two highly inbreeding populations of *Leavenworthia*. *Proc. R. Soc. B* **258**, 209–214. (doi:10.1098/rspb.1994.0164)
- Charlesworth, B., Borthwick, H., Bartolomé, C. & Pignatelli, P. 2004 Estimates of the genomic mutation rate for detrimental alleles in *Drosophila melanogaster*. *Genetics* **167**, 815–826. (doi:10.1534/genetics.103.025262)
- Chen, J. M., Chuzhanova, N., Stenson, P. D., Férec, C. & Cooper, D. N. 2005 Intrachromosomal serial replication slippage *in trans* gives rise to diverse genomic rearrangements involving inversions. *Hum. Mutat.* **26**, 362–373. (doi:10.1002/humu.20230)
- Crow, J. F. 1993 How much do we know about spontaneous human mutation rates? *Environ. Mol. Mutagen* **21**, 122–129. (doi:10.1002/em.2850210205)
- Crow, J. F. 2000 The origins, patterns and implications of human spontaneous mutation. *Nat. Rev. Genet.* **1**, 40–47. (doi:10.1038/35049558)
- Crow, J. F. 2006 Age and sex effects on human mutation rates: an old problem with new complexities. *J. Radiat. Res. (Tokyo)* **47**(Suppl. B), B75–B82.
- Crow, J. F. & Abrahamson, S. 1997 Seventy years ago: mutation becomes experimental. *Genetics* **147**, 1491–1496.
- Cutter, A. D. & Payseur, B. A. 2003 Rates of deleterious mutation and the evolution of sex in *Caenorhabditis*. *J. Evol. Biol.* **16**, 812–822. (doi:10.1046/j.1420-9101.2003.00596.x)
- Danforth, C. H. 1923 The frequency of mutation and the incidence of hereditary traits in man. In *Eugenics, genetics and the family*. Scientific papers of the 2nd International Congress of Eugenics, New York, 1921, vol. 1, pp. 120–128. Baltimore, MD: Williams & Wilkins.
- Deng, H. W. & Lynch, M. 1996 Estimation of deleterious-mutation parameters in natural populations. *Genetics* **144**, 349–360.
- Deng, H. W., Li, J., Pfreder, M. E., Li, J. L. & Deng, H. 2006 Upper limit of the rate and per generation effects of deleterious genomic mutations. *Genet. Res.* **88**, 57–65. (doi:10.1017/S0016672306008329)
- Denver, D. R., Morris, K., Lynch, M., Vassilieva, L. L. & Thomas, W. K. 2000 High direct estimate of the mutation rate in the mitochondrial genome of *Caenorhabditis elegans*. *Science* **289**, 2342–2344. (doi:10.1126/science.289.5488.2342)
- Denver, D. R., Morris, K., Lynch, M. & Thomas, W. K. 2004 High mutation rate and predominance of insertions in the *Caenorhabditis elegans* nuclear genome. *Nature* **430**, 679–682. (doi:10.1038/nature02697)
- Drake, J. W. 1993 Rates of spontaneous mutation among RNA viruses. *Proc. Natl Acad. Sci. USA* **90**, 4171–4175. (doi:10.1073/pnas.90.9.4171)
- Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. 1998 Rates of spontaneous mutation. *Genetics* **148**, 1667–1686.
- Drost, J. B. & Lee, W. R. 1995 Biological basis of germline mutation: comparisons of spontaneous germline mutation rates among *Drosophila*, mouse, and human. *Environ. Mol. Mutagen* **25**(Suppl. 26), 48–64.
- Eyre-Walker, A. & Keightley, P. D. 1999 High genomic deleterious mutation rates in hominids. *Nature* **397**, 344–347. (doi:10.1038/16915)
- Fernández, J. & López-Fanjul, C. 1996 Spontaneous mutational variances and covariances for fitness-related traits in *Drosophila melanogaster*. *Genetics* **143**, 829–837.
- Fry, J. D. 2001 Rapid mutational declines of viability in *Drosophila*. *Genet. Res.* **77**, 53–60. (doi:10.1017/S0016672300004882)
- Fu, Y. X. & Huai, H. 2003 Estimating mutation rate: how to count mutations? *Genetics* **164**, 797–805.
- Gago, S., Elena, S. F., Flores, R. & Sanjuán, R. 2009 Extremely high mutation rate of a hammerhead viroid. *Science* **323**, 1308. (doi:10.1126/science.1169202)
- Gong, Y., Gu, S. & Woodruff, R. C. 2005a The influence of premeiotic clusters of mutation on indirect estimations of mutation rate. *Hum. Hered.* **60**, 150–155. (doi:10.1159/000089955)
- Gong, Y., Woodruff, R. C. & Thompson Jr, J. N. 2005b Deleterious genomic mutation rate for viability in *Drosophila melanogaster* using concomitant sibling controls. *Biol. Lett.* **1**, 492–495. (doi:10.1098/rsbl.2005.0364)
- Gordenin, D. A. & Resnick, M. A. 1998 Yeast ARMs (DNA at-risk motifs) can reveal sources of genome instability. *Mutat. Res.* **400**, 45–58. (doi:10.1016/S0027-5107(98)00047-5)
- Green, P., Ewing, B., Miller, W., Thomas, P. J., Green, E. D. & NISC Comparative Sequencing Program 2003 Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet.* **33**, 514–517. (doi:10.1038/ng1103)
- Haag-Liautard, C., Dorris, M., Maside, X., Macaskill, S., Halligan, D. L., Houle, D., Charlesworth, B. & Keightley, P. D. 2007 Direct estimation of per nucleotide and genomic deleterious mutation rates in *Drosophila*. *Nature* **445**, 82–85. (doi:10.1038/nature05388)
- Haag-Liautard, C., Coffey, N., Houle, D., Lynch, M., Charlesworth, B. & Keightley, P. D. 2008 Direct estimation of the mitochondrial DNA mutation rate in *Drosophila melanogaster*. *PLoS Biol.* **6**, e204. (doi:10.1371/journal.pbio.0060204)
- Haldane, J. B. S. 1935 The rate of spontaneous mutation in a human gene. *J. Genet.* **31**, 317–326. (doi:10.1007/BF02982403)
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M. & Ira, G. 2009 Mechanisms of change in gene copy number. *Nat. Rev. Genet.* **10**, 551–564. (doi:10.1038/nrg2593)
- HAZARD DB 2009 <http://www.evol.nw.ru/~spirov/hazard/hazard.html>.
- Henn, B. M., Gignoux, C. R., Feldman, M. W. & Mountain, J. L. 2009 Characterizing the time dependency of human

- mitochondrial DNA mutation rate estimates. *Mol. Biol. Evol.* **26**, 217–230. (doi:10.1093/molbev/msn244)
- Hodgkinson, A., Ladoukakis, E. & Eyre-Walker, A. 2009 Cryptic variation in the human mutation rate. *PLoS Biol.* **7**, e27. (doi:10.1371/journal.pbio.1000027)
- Holliday, R. & Grigg, G. W. 1993 DNA methylation and mutation. *Mutat. Res.* **285**, 61–67.
- Houle, D. & Nuzhdin, S. V. 2004 Mutation accumulation and the effect of copia insertions in *Drosophila melanogaster*. *Genet. Res.* **83**, 7–18. (doi:10.1017/S0016672303006505)
- Howell, N., Smejkal, C. B., Mackey, D. A., Chinnery, P. F., Turnbull, D. M. & Herrnstadt, C. 2003 The pedigree rate of sequence divergence in the human mitochondrial genome: there is a difference between phylogenetic and pedigree rates. *Am. J. Hum. Genet.* **72**, 659–670. (doi:10.1086/368264)
- Keightley, P. D. & Caballero, A. 1997 Genomic mutation rates for lifetime reproductive output and lifespan in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **94**, 3823–3827. (doi:10.1073/pnas.94.8.3823)
- Keightley, P. D. & Eyre-Walker, A. 1999 Terumi Mukai and the riddle of deleterious mutation rates. *Genetics* **153**, 515–523.
- Keightley, P. D. & Eyre-Walker, A. 2000 Deleterious mutations and the evolution of sex. *Science* **290**, 331–333. (doi:10.1126/science.290.5490.331)
- Keightley, P. D. & Eyre-Walker, A. 2010 What can we learn about the distribution of fitness effects of new mutations from DNA sequence data? *Phil. Trans. R. Soc. B* **365**, 1187–1193. (doi:10.1098/rstb.2009.0266)
- Keightley, P. D. & Halligan, D. L. 2009 Analysis and implications of mutational variation. *Genetica* **136**, 359–369. (doi:10.1007/s10709-008-9304-4)
- Keightley, P. D. & Ohnishi, O. 1998 EMS-induced polygenic mutation rates for nine quantitative characters in *Drosophila melanogaster*. *Genetics* **148**, 753–766.
- Keightley, P. D., Trivedi, U., Thomson, M., Oliver, F., Kumar, S. & Blaxter, M. L. 2009 Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome Res.* **19**, 1195–1201. (doi:10.1101/gr.091231.109)
- Keller, I., Bensasson, D. & Nichols, R. A. 2007 Transition-transversion bias is not universal: a counter example from grasshopper pseudogenes. *PLoS Genet.* **3**, e22. (doi:10.1371/journal.pgen.0030022)
- Kibota, T. T. & Lynch, M. 1996 Estimate of the genomic mutation rate deleterious to overall fitness in *E. coli*. *Nature* **381**, 694–696. (doi:10.1038/381694a0)
- Kimura, M. 1968 Evolutionary rate at the molecular level. *Nature* **217**, 624–626. (doi:10.1038/217624a0)
- Kondrashov, A. S. 1998 Measuring spontaneous deleterious mutation process. *Genetica* **102–103**, 183–197. (doi:10.1023/A:1017085731998)
- Kondrashov, A. S. 2001 Sex and U. *Trends Genet.* **17**, 75–77. (doi:10.1016/S0168-9525(00)02188-0)
- Kondrashov, A. S. 2003 Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* **21**, 12–27. (doi:10.1002/humu.10147)
- Kondrashov, A. S. 2008 Another step toward quantifying spontaneous mutation. *Proc. Natl Acad. Sci. USA* **105**, 9133–9134. (doi:10.1073/pnas.0804749105)
- Kondrashov, A. S. & Crow, J. F. 1993 A molecular approach to estimating the human deleterious mutation rate. *Hum. Mutat.* **2**, 229–234. (doi:10.1002/humu.1380020312)
- Kondrashov, A. S. & Houle, D. 1994 Genotype–environment interactions and the estimation of the genomic mutation rate in *Drosophila melanogaster*. *Proc. Biol. Sci.* **258**, 221–227. (doi:10.1098/rspb.1994.0166)
- Kondrashov, A. S. & Rogozin, I. B. 2004 Context of deletions and insertions in human coding sequences. *Hum. Mutat.* **23**, 177–185. (doi:10.1002/humu.10312)
- Kondrashov, F. A., Ogurtsov, A. Y. & Kondrashov, A. S. 2006 Selection in favor of nucleotides G and C diversifies evolution rates and levels of polymorphism at mammalian synonymous sites. *J. Theor. Biol.* **240**, 616–626. (doi:10.1016/j.jtbi.2005.10.020)
- Kumar, S. & Subramanian, S. 2002 Mutation rates in mammalian genomes. *Proc. Natl Acad. Sci. USA* **99**, 803–808. (doi:10.1073/pnas.022629899)
- Lewin, B. 2008 *Genes IX*. Sudbury, MA: Jones and Bartlett.
- Li, W. H. 1997 *Molecular evolution*. Sunderland, MA: Sinauer.
- Li, J. & Deng, H. W. 2005 Estimation of the rate and effects of deleterious genomic mutations in finite populations with linkage disequilibrium. *Heredity* **95**, 59–68. (doi:10.1038/sj.hdy.6800689)
- Loewe, L., Textor, V. & Scherer, S. 2003 High deleterious genomic mutation rate in stationary phase of *Escherichia coli*. *Science* **302**, 1558–1560. (doi:10.1126/science.1087911)
- Lynch, M. 2006 The origins of eukaryotic gene structure. *Mol. Biol. Evol.* **23**, 450–468. (doi:10.1093/molbev/msj050)
- Lynch, M. 2008 The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* **180**, 933–943. (doi:10.1534/genetics.108.090456)
- Lynch, M., Blanchard, J., Houle, D., Kibota, T., Schultz, S., Vassilieva, L. & Willis, J. 1999 Perspective: spontaneous deleterious mutation. *Evolution* **53**, 645–663. (doi:10.2307/2640707)
- Lynch, M., Koskella, B. & Schaack, S. 2006 Mutation pressure and the evolution of organelle genomic architecture. *Science* **311**, 1727–1730. (doi:10.1126/science.1118884)
- Lynch, M. et al. 2008 A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl Acad. Sci. USA* **105**, 9272–9277. (doi:10.1073/pnas.0803466105)
- Mackwan, R. R., Carver, G. T., Kissling, G. E., Drake, J. W. & Grogan, D. W. 2008 The rate and character of spontaneous mutation in *Thermus thermophilus*. *Genetics* **180**, 17–25. (doi:10.1534/genetics.108.089086)
- Makova, K. D. & Li, W. H. 2002 Strong male-driven evolution of DNA sequences in humans and apes. *Nature* **416**, 624–626. (doi:10.1038/416624a)
- Malpica, J. M., Fraile, A., Moreno, I., Obies, C. I., Drake, J. W. & García-Arenal, F. 2002 The rate and character of spontaneous mutation in an RNA virus. *Genetics* **162**, 1505–1511.
- Mardis, E. R. 2006 Anticipating the 1,000 dollar genome. *Genome Biol.* **7**, 112. (doi:10.1186/gb-2006-7-7-112)
- Messer, P. W. 2009 Measuring the rates of spontaneous mutation from deep and large-scale polymorphism data. *Genetics* **182**, 1219–1232. (doi:10.1534/genetics.109.105692)
- Mitchell, A. & Graur, D. 2005 Inferring the pattern of spontaneous mutation from the pattern of substitution in unitary pseudogenes of *Mycobacterium leprae* and a comparison of mutation patterns among distantly related organisms. *J. Mol. Evol.* **61**, 795–803. (doi:10.1007/s00239-004-0235-0)
- Morton, R. A. & Morton, B. R. 2007 Separating the effects of mutation and selection in producing DNA skew in bacterial chromosomes. *BMC Genom.* **8**, 369. (doi:10.1186/1471-2164-8-369)
- Mukai, T. 1964 The genetic structure of natural populations of *Drosophila melanogaster*. I. Spontaneous mutation rate of polygenes controlling viability. *Genetics* **50**, 1–19.

- Mukai, T. 1970 Viability mutations induced by ethyl methanesulfonate in *Drosophila melanogaster*. *Genetics* **65**, 335–348.
- Mukai, T., Chigusa, S. I., Mettler, L. E. & Crow, J. F. 1972 Mutation rate and dominance of genes affecting viability in *Drosophila melanogaster*. *Genetics* **72**, 333–355.
- Muller, H. J. 1927 Artificial transmutation of the gene. *Science* **66**, 84–87. (doi:10.1126/science.66.1699.84)
- Muller, H. J. 1932 Further studies on the nature and causes of gene mutations. *Proc. Sixth Int. Congr. Genetics*, vol. 1, pp. 213–255.
- Nachman, M. W. 2004 Haldane and the first estimates of the human mutation rate. *J. Genet.* **83**, 231–233. (doi:10.1007/BF02717891)
- Nachman, M. W. & Crowell, S. L. 2000 Estimate of the mutation rate per nucleotide in humans. *Genetics* **156**, 297–304.
- Nuzhdin, S. V. & Mackay, T. F. 1994 Direct determination of retrotransposon transposition rates in *Drosophila melanogaster*. *Genet. Res.* **63**, 139–144. (doi:10.1017/S0016672300032249)
- Ohnishi, O. 1977 Spontaneous and ethyl methanesulfonate-induced mutations controlling viability in *Drosophila melanogaster* II. Homozygous effect of polygenic mutations. *Genetics* **87**, 529–545.
- Orr, H. A. 2010 The population genetics of beneficial mutations. *Phil. Trans. R. Soc. B* **365**, 1195–1201. (doi:10.1098/rstb.2009.0282)
- Pavlov, Y. I., Shcherbakova, P. V. & Rogozin, I. B. 2006 Roles of DNA polymerases in replication, repair, and recombination in eukaryotes. *Int. Rev. Cytol.* **255**, 41–132. (doi:10.1016/S0074-7696(06)55002-8)
- Petrov, D. A., Sangster, T. A., Johnston, J. S., Hartl, D. L. & Shaw, K. L. 2000 Evidence for DNA loss as a determinant of genome size. *Science* **287**, 1060–1062. (doi:10.1126/science.287.5455.1060)
- Roles, A. J. & Conner, J. K. 2008 Fitness effects of mutation accumulation in a natural outbred population of wild radish (*Raphanus raphanistrum*): comparison of field and greenhouse environments. *Evolution* **62**, 1066–1075. (doi:10.1111/j.1558-5646.2008.00354.x)
- Rogozin, I. B. & Pavlov, Y. I. 2003 Theoretical analysis of mutation hotspots and their DNA sequence context specificity. *Mutat. Res.* **544**, 65–85.
- Schrag, S. J., Rota, P. A. & Bellini, W. J. 1999 Spontaneous mutation rate of measles virus: direct estimation based on mutations conferring monoclonal antibody resistance. *J. Virol.* **73**, 51–54.
- Schultz, S. T., Lynch, M. & Willis, J. H. 1999 Spontaneous deleterious mutation in *Arabidopsis thaliana*. *Proc. Natl Acad. Sci. USA* **96**, 11 393–11 398. (doi:10.1073/pnas.96.20.11393)
- Shabalina, S. A., Yampolsky, L. Y. u. & Kondrashov, A. S. 1997 Rapid decline of fitness in panmictic populations of *Drosophila melanogaster* maintained under relaxed natural selection. *Proc. Natl Acad. Sci. USA* **94**, 13 034–13 039. (doi:10.1073/pnas.94.24.13034)
- Silva, J. C. & Kondrashov, A. S. 2002 Patterns in spontaneous mutation revealed by human–baboon sequence comparison. *Trends Genet.* **18**, 544–547. (doi:10.1016/S0168-9525(02)02757-9)
- Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. 1997 Evolution of high mutation rates in experimental populations of *E. coli*. *Nature* **387**, 703–705. (doi:10.1038/42701)
- Sommer, S. S. 1995 Recent human germ-line mutation: inferences from patients with hemophilia B. *Trends Genet.* **11**, 141–147. (doi:10.1016/S0168-9525(00)89028-9)
- Sommer, S. S. & Ketterling, R. P. 1996 The factor IX gene as a model for analysis of human germline mutations: an update. *Hum. Mol. Genet.* **5**, 1505–1514.
- Tanaka, M. & Ozawa, T. 1994 Strand asymmetry in human mitochondrial DNA mutations. *Genomics* **22**, 327–335. (doi:10.1006/geno.1994.1391)
- Timofeeff-Ressovsky, N. W. 1935 Auslösung von Vitilismutationen durch Röntgenbestrahlung bei *Drosophila melanogaster*. *Nachr. Gess. Wiss. Göttingen. Biol. N. F.* **1**, 163–180.
- Trindade, S., Perfeito, L. & Gordo, I. 2010 Rate and effects of spontaneous mutations that affect fitness in mutator *Escherichia coli*. *Phil. Trans. R. Soc. B* **365**, 1177–1186. (doi:10.1098/rstb.2009.0287)
- Unniraman, S. & Schatz, D. G. 2007 Strand-biased spreading of mutations during somatic hypermutation. *Science* **317**, 1227–1230. (doi:10.1126/science.1145065)
- Vassilieva, L. L. & Lynch, M. 1999 The rate of spontaneous mutation for life-history traits in *Caenorhabditis elegans*. *Genetics* **151**, 119–129.
- Wyman, C. & Kanaar, R. 2006 DNA double-strand break repair: all's well that ends well. *Annu. Rev. Genet.* **40**, 363–383. (doi:10.1146/annurev.genet.40.110405.090451)
- Xue, Y. *et al.* 2009 Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* **19**, 1453–1457. (doi:10.1016/j.cub.2009.07.032)
- Yang, H. P., Tanikawa, A. Y. & Kondrashov, A. S. 2001a Molecular nature of 11 spontaneous de novo mutations in *Drosophila melanogaster*. *Genetics* **157**, 1285–1292.
- Yang, H. P., Tanikawa, A. Y., Van Voorhies, W. A., Silva, J. C. & Kondrashov, A. S. 2001b Whole-genome effects of ethyl methanesulfonate-induced mutation on nine quantitative traits in outbred *Drosophila melanogaster*. *Genetics* **157**, 1257–1265.