Review in Advance first posted online on June 5, 2014. (Changes may still occur before final publication online and in print.)

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org by University of Winnipeg on 08/28/14. For personal use only.

Determinants of Mutation Rate Variation in the Human Germline

Laure Ségurel,[1,*] Minyoung J. Wyman,[2,*] and Molly Przeworski[3,**]

[1]Laboratoire Éco-Anthropologie et Ethnobiologie, UMR 7206, Muséum National d'Histoire Naturelle–Centre National de la Recherche Scientifique–Université Paris 7 Diderot, Paris 75231, France; email: lsegurel@mnhn.fr

[2]Department of Biological Sciences, Columbia University, New York, NY 10027; email: mw2930@columbia.edu

[3]Department of Human Genetics and Howard Hughes Medical Institute, University of Chicago, Chicago, Illinois 60637; email: mp3284@columbia.edu

Annu. Rev. Genomics Hum. Genet. 2014. 15:19.1–19.24

The *Annual Review of Genomics and Human Genetics* is online at genom.annualreviews.org

This article's doi: 10.1146/annurev-genom-031714-125740

Copyright © 2014 by Annual Reviews. All rights reserved

*These authors contributed equally to this work.

**Present address: Department of Biological Sciences and Department of Systems Biology, Columbia University, New York, NY 10027

Keywords

germline mutation, human, CpG, paternal age effect, generation time effect, evolution

Abstract

Because germline mutations are the source of all evolutionary adaptations and heritable diseases, characterizing their properties and the rate at which they arise across individuals is of fundamental importance for human genetics. After decades during which estimates were based on indirect approaches, notably on inferences from evolutionary patterns, it is now feasible to count de novo mutations in transmissions from parents to offspring. Surprisingly, this direct approach yields a mutation rate that is twofold lower than previous estimates, calling into question our understanding of the chronology of human evolution and raising the possibility that mutation rates have evolved relatively rapidly. Here, we bring together insights from studies of human genetics and molecular evolution, focusing on where they conflict and what the discrepancies tell us about important open questions. We begin by outlining various methods for studying the properties of mutations in humans. We review what we have learned from their applications about genomic factors that influence mutation rates and the effects of sex, age, and other sources of interindividual variation. We then consider the mutation rate as a product of evolution and discuss how and why it may have changed over time in primates.

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

INTRODUCTION

Germline mutations—the random changes to the genome that occur during the formation of egg and sperm—underlie all evolutionary adaptations and all heritable diseases. Thus, understanding the mechanisms by which they arise and estimating their rate of occurrence is of fundamental importance, whether the aim is to scan for de novo disease mutations (e.g., 103), interpret the paternal age effect on specific diseases (8, 28), make inferences about natural selection (e.g., 101, 144), or develop plausible theories about the evolution of sex (reviewed, e.g., in 10). Moreover, an accurate estimate of the mutation rate is essential to dating and interpreting demographic events inferred from genetic data (e.g., 133, 142) because the number of mutations that have accumulated between two chromosomes provides a record of the time elapsed since their common ancestor—a "molecular clock" (76).

In humans, where experimental approaches are obviously not feasible, estimates have relied primarily on the prevalence of monogenic diseases or on phylogenetic distances among primates (24, 55, 79, 96, 111). These methods have recently been upended by the advent of next-generation sequencing, which in principle makes a much more direct and comprehensive approach feasible: counting the germline point mutations that arise in transmissions from parents to offspring (19, 25, 46, 70, 71, 80, 103, 115, 119, 137, 141, 175).

Surprisingly, these pedigree studies yield estimates of the mutation rate that disagree with earlier approaches. The discrepancies among estimators cast considerable uncertainty over the chronology of human evolution and raise the possibility that the mutation rate may have changed relatively quickly across primates (142). Moreover, they highlight large gaps in our understanding of how mutations are generated, notably about the importance of replication errors in the genesis of mutations.

Motivated by these considerations, we review studies of germline mutations from human genetics and molecular evolution, with particular attention to where their results conflict. We begin with a brief description of various approaches to studying mutation rates and properties in humans. We consider what we have learned from their applications about the determinants of point mutations, the average mutation rate, and sources of interindividual variation. In so doing, we discuss the well-known discrepancies between mutation rate estimates obtained from different methods but also highlight unappreciated incongruities among pedigree studies. Finally, we describe the selective pressures to which mutation rates may be subject, either directly or indirectly, and how rates may have evolved in primates.

Throughout, our primary focus is on germline point mutations in humans. For information on other organisms or on the many other types of mutations that arise through distinct mechanisms, we refer readers to other recent reviews (10, 20, 31, 93, 95). Furthermore, we do not cover the growing literature on somatic mutations (e.g., 6, 84, 131), other than to point out salient properties shared with the germline.

HOW WE LEARN ABOUT GERMLINE MUTATIONS IN HUMANS

Our understanding of how DNA polymerases synthesize DNA and how errors occur in the replication process comes from biochemical studies (reviewed in 37, 52) that compare the behavior of various mutated polymerase alleles (12), assess the influence of inserting base analogs during replication (13, 108, 158), or measure transition and transversion mutation rates for polymerases in different sequence contexts through enzyme kinetic analyses (51). These studies cannot address sources of mutations in their native, in vivo environment, however, and in particular they do not provide estimates of germline mutation rates.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

To address these questions in nonmodel organisms such as humans, one of four approaches has been taken. Historically, the first estimates in humans were based on the hypothesis that severe Mendelian diseases arise from a balance between mutation and natural selection in the population, such that their frequency is expected to be approximately u/hs for dominant disorders (where u is the rate at which mutations give rise to a given disease and hs is the fitness decrease in heterozygotes). The mutation rate per base pair can then be estimated from the frequency of the disease, given estimates of hs and the mutational target size (i.e., the number of sites at which mutations would lead to the disease) (55, 79, 96, 110). These disease incidence–based approaches make several assumptions, not only about the target size, obviously, but also in requiring complete ascertainment of the phenotype. Moreover, the overall mutation rate estimate may not be representative of the genome, because it tends to be based on exons, and specific types of mutations (e.g., changes at CpG sites) may be overrepresented among disease alleles (26).

A second, phylogenetic approach relies on the fact that at neutral sites, the rate at which substitutions accumulate between species is equal to the mutation rate per generation (76). Thus, the number of substitutions at putatively neutral sites (e.g., pseudogenes) can be converted into a mutation rate over millions of years, given an independent estimate of the species split time (usually a date in years, from securely dated fossils). This conversion depends on an estimate of the generation time and on a calibration point, and so is fraught with uncertainty. Moreover, if nonneutral sites are included, the mutation rate will be underestimated because deleterious alleles will tend to be purged by natural selection and not contribute to divergence. Even at neutral sites, processes such as GC-biased gene conversion [i.e., the preferential repair of A/T-G/C mismatches that arise during recombination to G-C (120)], which act analogously to selection (112), may lead the substitution rate to differ from the mutation rate. An additional difficulty is that comparisons among closely related species, such as humans and chimpanzees, strongly rely on assumptions about the extent to which polymorphism in the ancestral species contributes to the divergence between them. In turn, comparisons among more distant species require a reliable correction for multiple substitutions at a site.

Neutral divergence between species can also be used to estimate the ratio of mutations that arise in the male germline versus the female germline (α) by contrasting patterns on autosomes, X chromosomes, and Y chromosomes (97, 104). Underlying this approach is the fact that, assuming an equal sex ratio, autosomes spend an equal number of generations in males and females, whereas X chromosomes are present in males only one-third of the time and Y chromosomes are always carried by males. Thus, if most mutations are introduced through males, the Y should show higher neutral divergence than autosomes, which should show higher divergence than the X, and the magnitude of the differences provides an estimate of the male bias. An important complication is that comparisons among closely related species (such as humans and chimpanzees) are highly sensitive to assumptions about the contribution of ancestral polymorphism to divergence, because levels of ancestral polymorphism are expected to differ between the autosomes, X, and Y, in ways that depend on the effects of linked selection (101) and on demographic history. In addition, this approach relies on the assumptions that the mutation rate in each sex is the same for the X and the autosomes (100) and that the degree of male bias has not changed along a lineage.

A third method consists of analyzing genetic variation data from multiple humans. Because neutral polymorphism levels reflect the rates at which mutations arise, variation data can be used to estimate relative mutation rates for different types of sites (in principle, absolute rates can also be estimated by assuming an effective population size) (76). A difficulty is that polymorphism levels will reflect not only the mutation process but also the effects of both biased gene conversion (35) and—unless sites are strictly neutral—natural selection, which may differ across types of sites. This limitation can be partially overcome by considering huge samples (i.e., of tens of thousands

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

of individuals), which are now available for exomes and a subset of genomic regions (e.g., 144, 156) and in which low-frequency variants more faithfully reflect the mutational spectrum.

A more direct approach, which has only recently become feasible, is to survey most of the genome, or a small subset of the genome (e.g., exomes), in pedigrees and identify mutations as variants present in children but not their parents (118, 137). This pedigree-based method provides an estimate of the mutation rate in each gamete and can be used to quantify variation among parents with regard to parameters of interest (age, disease status, etc.). It can be applied to DNA extracted from blood or other tissues or by sequencing DNA from single sperm (or, in principle, oocytes) (162). Through experimental or bioinformatic approaches [e.g., phasing using additional single-nucleotide polymorphisms (SNPs) in the sequencing reads] or sequencing of a third generation, mutations can be ascribed to maternal and paternal origins (25, 80).

In principle, the pedigree-based approach offers a comprehensive and relatively unbiased picture of the rate at which different types of mutation arise. In practice, the task amounts to looking for a needle in a haystack of 100 million base pairs. Numerous filters have to be applied to weed out false positives, e.g., by excluding heterozygous positions with an allelic imbalance in coverage or sites at which a mutation is already known in databases (i.e., discounting the possibility of more than one mutation among the sampled individuals at the same site). Because these filters also remove a subset of real mutations, obtaining an accurate count of the number of de novo mutations requires a reliable estimate of the false-negative rate (as well as the false-discovery rate). Perhaps even more problematic is obtaining a reliable estimate of the denominator for the mutation rate, i.e., the number of sites at which mutations would have been found with high power had they occurred. This challenge is, in our view, underappreciated, with some studies considering more of the genome in the denominator than is readily accessible by current sequencing technologies. Another issue that has received relatively little attention to date is the effect of somatic mutations in the offspring when using blood (or other nongermline) samples. Although in principle the contribution of somatic mutations can be assessed by sequencing a third generation, in practice this has been done in only a small subset of cases (25, 80). Thus, what fraction of each type of mutation is somatic in origin remains unclear.

HOW GERMLINE MUTATIONS ARISE

Sources of Mutations

Mutations arise by various mechanisms that can be grouped into two categories: copying errors that arise during replication, and DNA changes that occur spontaneously (i.e., nonreplicatively) because of the instability of the structure of DNA (91) or are induced by endogenous or exogenous sources (i.e., radiation or chemical and biological agents). Both types of errors lead to mutations if they are not repaired correctly before the next round of replication. The number of mutations thus represents a balance between the number of errors introduced and the efficacy of various types of repair. One implication is that mutation rates could differ between individuals and sexes because of differences in error rates (32) or repair efficiencies (105).

Replication errors arise from incorrect base incorporation (resulting in SNPs) or slippage of the polymerase (resulting in small insertions or deletions, typically <20 base pairs) (138). One important factor influencing replication fidelity is the neighboring base composition. In particular, some DNA polymerases also have exonuclease activity, which allows them to excise misincorporated nucleotides and therefore increase fidelity (17, 109). This repair process is facilitated in regions where DNA strands can be easily separated: Because A-T base pairs (double-bonded adenine and thymine) are connected by only two hydrogen bonds, whereas G-C base pairs (triple-bonded

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

guanine and cytosine) are connected by three, DNA in AT-rich regions is easier to separate, leading to more effective repair in these regions by the exonuclease domain than in GC-rich regions (15, 127). This phenomenon could be responsible for the higher mutation rate in GC-rich regions (58).

Errors can also occur because of spontaneous or induced nucleotide changes. Remarkably, the mammalian nuclear genome is thought to contain a steady state of 50,000–200,000 apurinic or apyrimidinic sites (DNA sites that have lost their purine or pyrimidine base but still have an intact backbone) per somatic cell, which are created by endogenous lesions (113) and are constantly being repaired in a manner dependent on the source of the DNA damage.

In addition, C can spontaneously undergo hydrolytic deamination to uracil (92). In mammals, the 5′ C in a CpG context is usually methylated (14, 39), and methylated C undergoes spontaneous deamination to T (27). Whereas the presence of uracil is easily recognized by the uracil glycosylase and efficiently repaired by the base excision pathway, T-G mismatches are repaired by less efficient repair pathways (47, 146), resulting in a higher prevalence of C-to-T transitions at methylated CpG sites (27, 34). Not all CpG transitions necessarily derive from this mechanism, however (128). Moreover, transversions also occur at a higher rate at CpG sites compared with other dinucleotides (79, 111, 148), for reasons that are still unclear. One possibility is that the presence of a methyl group at CpG sites enhances adduct formation [the bonding of DNA with other molecules (174)] at the adjacent G, as shown, e.g., for benzo(*a*)pyrene (174), a mutagenic component of tobacco smoke that preferentially induces G-to-T transversions (128). Another example of spontaneous mutation is the G-to-T transversion that results from the insertion of 8-oxoguanine, an endogenous molecule (121).

Beyond these endogenous sources, recent studies of tumors have led to a better understanding of the effects of environmental mutagens. For instance, skin cancers bear C-to-T and CC-to-TT transitions at dipyrimidine sequences (131) as a result of UV light exposure (129), and lung tumors demonstrate frequent G-to-T transversions (132) caused by benzo(*a*)pyrene (140). Smoking-related DNA adducts are also found in oocytes and spermatozoa (reviewed in 176), suggesting that it would be interesting to look for similar mutational signatures among germline mutations.

Mutation Rate Variation Along the Genome

Mutation rates vary along the genome at both fine and broad scales (12). Considering adjacent base pairs, the strongest effect is an elevated mutation rate at CpG sites (67, 69). Even though CpGs are rare in mammalian genomes, constituting less than 2% of sites in the human genome, they account for ∼19% of de novo mutations. Notably, transitions at CpGs are 12–13-fold more likely than transitions at non-CpG bases in polymorphism and divergence data (86, 111), 15-fold more likely in disease incidence studies (79), and 18-fold more likely in pedigree data (80). These numbers are consistent with the hypothesis that spontaneous deamination at CpGs frequently generates transitions from C to T (or G to A). In turn, the lower numbers in polymorphism and divergence compared with de novo mutations may reflect the effects of biased gene conversion favoring G or C alleles (21, 120) and purifying selection to maintain CpGs (145).

Across the genome, the rate of transitions is twofold higher than the rate of transversions (44, 80, 90, 144), whereas the opposite would be expected if all mutations were equally likely (because there are four possible transitions for eight possible transversions). This observation has at least two sources: First, CpG transitions are the most common mutation type. Second, DNA mispairing during replication occurs at varying frequencies, with G-T and A-C being the most common (37, 160) and resulting in transitions.

There is also a twofold higher rate of mutation from strong bases (G or C) to weak bases (A or T) compared with the reverse (80, 144), which may be due to the higher mutation rate at CpGs

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

and the decreased efficacy of the exonuclease in regions of high GC content (15). In spite of the decreased access, CpGs are less mutagenic in regions of high GC content (47), possibly because deamination of Cs depends on local strand separation, which occurs less often when there is strong hydrogen bonding in the neighboring region.

At an intermediate scale, additional, weaker context-dependent effects that currently lack a mechanistic explanation have also been reported (7, 66, 67, 69, 148), such as a higher mutation rate away from T with an increasing number of flanking purines (69). Low-complexity repetitive DNA may also be mutagenic: Controlling for adjacent base pairs, there is an excess of SNPs shared between humans and chimpanzees in regions with repetitive DNA (86), consistent with sequence-dependent replication slippage (18, 86).

Over broader scales, transcription and DNA replication timing are influential. Recent studies have revealed a higher mutation rate and SNP density in late-replicating regions in the germline (152, 163) as well as in tumors (84), possibly owing to a depletion of free nucleotides during replication (152). Moreover, mutational patterns are affected by transcription both in the germline (102, 131) and in tumors (84, 139). This is likely due in part to transcription-coupled repair (45): Because the nontranscribed (coding) strand is used as the template for repair, a characteristic pattern is expected, namely an asymmetry in base composition between coding and template strands (54). In humans, there is a 40% increase in the rate of A-to-G transitions in the nontranscribed strand compared with the transcribed strand (54, 154), and strand asymmetry is more pronounced for genes highly expressed in the germline (102). Increased gene expression levels are also associated with a decrease in the mutation rate in tumors (84, 131), but this effect is weak or absent in the germline (67).

Other factors have also been proposed to affect mutation rates. For example, studies of diversity and divergence raised the possibility that meiotic recombination introduces point mutations (62, 63, 87). In studies of de novo mutations, however, the effect looks to be at most very small and possibly noncausal (144, 152). Similarly, other factors such as nucleosome occupancy and DNase hypersensitivity are correlated with broad-scale mutational patterns (67, 103, 173), but the association could be due, at least in part, to confounders such as base composition.

Nonindependence of Mutations

Multiple mutations in the same gamete may not always be independent. Over distances on the order of 10 kb, mutations that arose in a single transmission appear to be more clustered in the genome than would be expected under a model where they occur independently (19, 60, 103, 147, 154). The molecular mechanism that gives rise to clustering on that scale is unknown.

Focusing on shorter distances of less than 100 base pairs, there is an excess of pairs of mutations in perfect association in population samples (i.e., in complete linkage disequilibrium), suggesting that they may often arise during the same replication cycle (59). Inspection of the pairs points to an error-prone polymerase used in translesion synthesis as the likely source (59, 153). Normal DNA polymerases cannot bypass some DNA lesions (e.g., oxidized bases and alkylated bases) during synthesis, requiring instead the intervention of special polymerases such as Pol zeta, which is error prone (153). Before the error-prone polymerase is switched out in favor of the high-fidelity polymerase, there is a higher chance of more than one mutation occurring.

In addition, mutations seen in more than one offspring may reflect a single mutation in the parental germline. Indeed, any mutation that arises before the final meiotic division in the germline lineage could in principle be shared among gametic stem cells, gametes, and/or offspring (168). Consistent with this expectation, Kong et al. (80) and Iossifov et al. (70) observed a shared de novo point mutation between siblings. A shared mutation has also been observed in the model organism

REVIEWS IN ADVANCE
Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

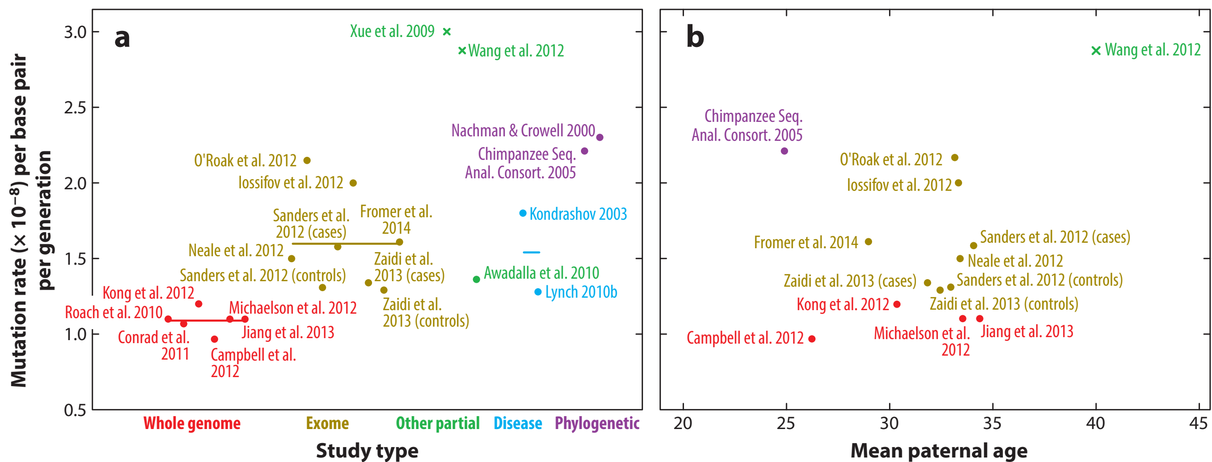

Figure 1

Estimates of the human mutation rate per base pair per generation. Plotted are estimates of (*a*) the human mutation rate estimated using different approaches and (*b*) the human mutation rate as a function of the mean paternal age for the studies where those data were available. Colored horizontal lines indicate the average mutation rate within each study type. We calculated a phylogenetic mutation rate based on a human–chimpanzee genetic divergence of 1.23% (23, 24, 36), a human–chimpanzee divergence time (i.e., time to the most recent common ancestor) of 7 million years, and a sex-averaged generation time of 25 years (43). The Nachman & Crowell (111) estimate is based on nucleotide substitutions only and is provided under their assumptions (a divergence time of 5.4 million years and a generation time of 20 years). To be consistent among studies, we calculated the paternal age at birth (by adding nine months as appropriate when studies reported the paternal age at conception). The Michaelson et al. (103) estimate has been revised to take into account the false-negative rate reported by the authors. The Jiang et al. (71) estimate was not reported by the authors; instead, we obtained it from their counts using the denominator and the false-positive and false-negative rates from Michaelson et al. (103), who had a very similar study design. Conrad et al. (25) measured the mutation rate in a European (CEU) trio and Yoruban (YRI) trio separately; we plotted the average. Sanders et al. (141) sequenced the exomes of autistic probands (cases) as well as unaffected siblings (controls). Iossifov et al. (70) sequenced the exomes of autistic probands and their unaffected siblings but reported the combined mutation rate (no significant difference was found). Zaidi et al. (175) sequenced individuals with congenital heart disease (cases) and unrelated individuals (controls). Mutation rates for exome studies were not adjusted for error rates, as the false-negative rate was not consistently reported and most studies validated all de novo mutations. "Other partial" refers to studies that sequenced targeted regions or other subsets of the genome. Xue et al. (171) sequenced the Y chromosomes of two individuals separated by 13 generations, and Wang et al. (162) sequenced eight sperm from a single individual. Kondrashov (79) and Lynch (96) calculated the mutation rate from disease incidences.

Drosophila melanogaster (72). More cases are expected as families of more than one offspring are sequenced.

THE AVERAGE MUTATION RATE AND INTERINDIVIDUAL VARIATION

The Number of De Novo Mutations Inherited by Humans

Whole genome pedigree-based estimates yield a mutation rate of 10^{-8} per base pair per generation. This is about two-thirds of the rate estimated by exome sequencing (**Figure 1*a***), consistent with a more than threefold enrichment of CpGs in exons compared with the rest of the genome as well as with their 20% higher GC content (115). Perplexingly, however, the mutation rate estimated by whole-genome pedigree studies is also more than twofold below the rates obtained from phylogenetic approaches (142) (**Figure 1*a***), even though it might be expected, if anything, to be slightly higher (because it includes deleterious mutations that would eventually be weeded out by selection).

REVIEWS IN ADVANCE
Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

Part of the difference between the substitution rates measured from phylogenetic data and the de novo mutations identified in pedigrees could, in theory, be due to biased gene conversion favoring the fixation of mutations to G or C. A recent study found that less than 2% of the divergence to chimpanzee was driven by biased gene conversion (21), however, so this is likely to be a small effect genome-wide.

A simpler explanation is that phylogenetic estimates are in error, either because the time to the most recent common ancestor of human and chimpanzee has been underestimated or because the generation time has been quite a bit lower than it is now (in either humans or chimpanzees) for most of hominoid evolution. For example, if the average time to the most recent common ancestor were 8 Mya and the generation time over much of that period hovered around 20, then the substitution rate might be as low as $0.0123/(2 \times 400{,}000 \text{ generations}) = 1.5 \times 10^{-8}$ per base pair per generation—only ∼50% higher than estimates based on whole-genome sequencing of pedigrees.

Because pedigree-based approaches do not rely on uncertain parameter estimates, it is tempting to consider their results to be the gold standard and to revise our estimate of the germline mutation rate downward. A closer look at the pedigree studies, however, suggests that this may be premature. First, although the six whole-genome studies differ in study design, analysis tools, and filters, the point estimates of the sex-averaged mutation rate do not seem to be entirely independent across studies, as they appear underdispersed: All fall between 0.96×10^{-8} and 1.20×10^{-8} per base pair per generation (**Figure 1*a***). Second, jointly considering the four whole-genome studies with age information (19, 71, 80, 103) would suggest little, if any, paternal age effect, and yet a marked effect of the father's age is seen in three of the whole-genome studies (**Figure 1*b***). Moreover, there appears to be a significant difference between the strength of the paternal age effect estimated by Kong et al. (80) and that from the combined data from the three other studies (19, 20, 71, 103) (**Figure 2**), which could be a true difference or could instead reflect the impact of study design or analysis choices.

Also troubling is that divergence patterns at mitochondrial sequences of securely dated fossils (using carbon dating) from the past 40,000 years are inconsistent with the low pedigree-based estimate of the mutation rate (48). Together, these findings lend uncertainty about the human germline mutation rate (for a given paternal age).

In interpreting the difference between pedigree-based and phylogenetic estimates of the mutation rate, it is worth noting that this difference is greater for transitions at non-CpGs (2.2-fold, using the data from References 80 and 111) than for transitions at CpGs (1.7-fold) or transversions at non-CpGs (1.65-fold). The differences among mutation types could be real or could point to technical issues, such as an effect of base composition on the power to detect de novo mutations or on the accuracy of substitution rate estimates. If real, these patterns imply that biased gene conversion cannot account for the discrepancy between pedigree-based and phylogenetic estimates, as transitions at CpGs will be undertransmitted in heterozygotes (i.e., "selected against" by biased gene conversion), and yet the substitution rate is higher than the pedigree-based estimate for this type of mutation as well. More generally, the differences suggest that, in making sense of the discrepancy among estimators, it may be informative to break down mutations by their likely mechanistic sources.

Effects of Sex and Age

In mammals, species with a shorter generation time tend to have higher neutral substitution rates per unit time. This so-called generation time effect is the basis for the hypothesis that most mutations arise from replication errors (50, 89, 122, 169) (see sidebar The Effect of Generation Time on the Mutation Rate Per Year). Also consistent with this hypothesis, the few direct

Changes may still occur before final publication online and in print

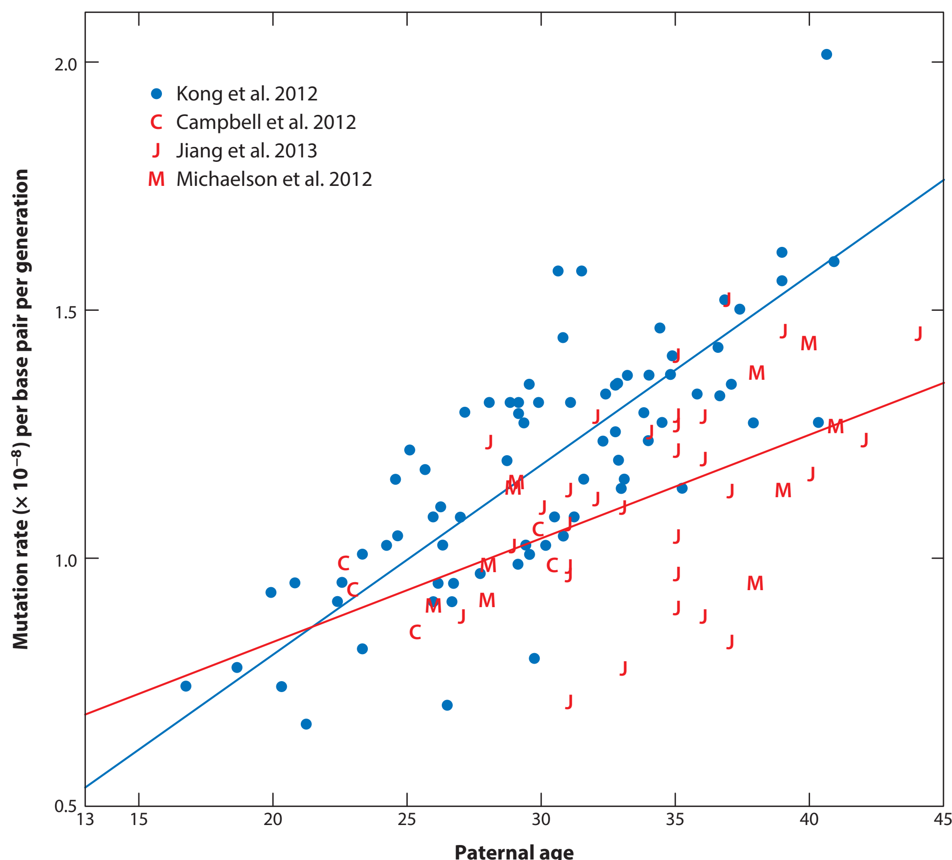

Figure 2

Differences among whole-genome pedigree studies. In an ordinary least squares regression of mutation rate on paternal age at birth, the trios from Kong et al. (80) (*blue dots*) have a significantly higher slope ($F_{1,121} = 8.98$, $P = 0.003$) than do trios from three smaller studies [C, Campbell et al. (19); J, Jiang et al. (71); M, Michaelson et al. (103)]. The slopes for the three smaller studies do not differ significantly from one another ($F_{2,41} = 0.02$, $P = 0.98$).

estimates of mutation rates per generation in eukaryotes suggest an increase with the number of cell divisions (**Figure 3**).

Replication-driven mutations. In humans, a preponderance of replication-driven mutation is supported by differences in male and female mutation patterns. In humans, oogonial mitotic divisions have formed the entire pool of primary oocytes by the birth of the future mother. Each primary oocyte has undergone approximately 31 cell divisions (33). In males, in contrast, before spermatogenesis, there are 34 cell divisions (33). After puberty, sperm are continuously produced through the asymmetric division of self-renewing spermatogonial stem cells every 16 days (61). Five additional cell divisions (corresponding to four replication events) are required to obtain mature sperm. As a consequence, if puberty occurs at 13 years (116), approximately $38 + 23 \times (30 - 13) = 429$ stem cell divisions have occurred if conception takes place at a paternal age of 30 years, and approximately 659 have occurred if conception takes place at a paternal age

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org by University of Winnipeg on 08/28/14. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

THE EFFECT OF GENERATION TIME ON THE MUTATION RATE PER YEAR

Mammalian species with longer generation times tend to have lower neutral substitution rates per year (50, 89, 122, 169). This pattern has been interpreted as evidence that most mutations are replication driven, assuming that increasing the generation time decreases the number of cell divisions per year (42, 89). Here, we derive the conditions for such an effect of generation time on mutation rates to hold.

The number of cell divisions per generation can be modeled as follows: Let B be the time from conception to birth, P the average age of puberty in males, and G the average age of reproduction in males. In females, all cell divisions are over by birth and can be approximated as occurring at a yearly rate c_0^{f} throughout that period. In males, the number of cell divisions, c, is the sum over three periods, where c_0^{m} is the yearly rate of cell division before birth, c_1 the yearly rate between birth and puberty, and c_2 the yearly rate after puberty.

All else being equal, we can therefore model the replication-driven mutation rate per generation for autosomes, m, as

$$m = (\mu_0^{\mathrm{m}} c_0^{\mathrm{m}} + \mu_0^{\mathrm{f}} c_0^{\mathrm{f}})B + \mu_1^{\mathrm{m}} c_1 (P - B) + \mu_2^{\mathrm{m}} c_2 (G - P), \tag{1}$$

where μ_0^{m} and μ_0^{f} are the male and female mutation rates per cell division between conception and birth, respectively, and μ_1^{m} and μ_2^{m} are the male mutation rates per cell division before and after puberty, respectively. Because in humans $c_1 \sim 0$, Equation 1 reduces to $(\mu_0^{\mathrm{m}} c_0^{\mathrm{m}} + \mu_0^{\mathrm{f}} c_0^{\mathrm{f}})B + \mu_2^{\mathrm{m}} c_2 (G - P)$, and the mutation rate per year, m_{y}, is then

$$m_{\mathrm{y}} = \mu_2^{\mathrm{m}} c_2 + [(\mu_0^{\mathrm{m}} c_0^{\mathrm{m}} + \mu_0^{\mathrm{f}} c_0^{\mathrm{f}})B - \mu_2^{\mathrm{m}} c_2 P]/G. \tag{2}$$

From Equation 2, it becomes clear that an increase in the generation time, G, should lead to a decrease in the mutation rate per year, m_{y}, if and only if $(\mu_0^{\mathrm{m}} c_0^{\mathrm{m}} + \mu_0^{\mathrm{f}} c_0^{\mathrm{f}})B/P > \mu_2^{\mathrm{m}} c_2$, assuming $c_1 = 0$. Also clear is that other ontogenetic changes in B or P (such as an increase in P) or in the yearly rate of cell division will affect m_{y} as well.

of 40 years. If mutations are due largely to replication errors, we therefore expect most to be introduced in the paternal germline, at a rate that depends on paternal age.

A greater contribution of paternal mutations was first noted in studies of human Mendelian disease (29, 55). This observation was lent strong support by molecular evolution studies of autosomes and X and Y chromosomes that reported lower divergence on the X than on the Y and autosomes (97, 104). The precise estimates of the ratio of male to female mutations (α) from phylogenetic data vary considerably among studies depending on which primates are compared with humans, what assumptions are made about the ancestral population size for the autosomes and X, and whether CpGs are excluded, but they center around 3 (133, 143, 155, 170) (see **Figure 4**). Pedigree studies likewise suggest that α is ~4 for all sites (for a mean paternal age of 30) (80) and ~3 for non-CpGs (for a paternal age of 30) (see **Figure 4**).

A paternal age effect on mutation rates was first deduced from the increased incidence of certain dominant diseases (e.g., Apert syndrome) in the children of older fathers (28, 56, 125, 164). For these disorders, however, the relationship between incidence rate and paternal age is often nonlinear (136), and recent studies have suggested that it could reflect a transmission advantage of the disease mutations within the germline (beyond the accumulation of replication-driven errors) (for recent reviews, see 8, 53). Beyond these specific cases, the prevalence of this type of transmission distortion remains unknown.

Considering all de novo mutations identified in pedigree studies, rather than only mutations associated with disease, there is also a clear increase in the number of mutations with paternal age. This increase appears to be close to linear after puberty, adding 1.5–2 mutations yearly (80)

REVIEWS IN ADVANCE
Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org by University of Winnipeg on 08/28/14. For personal use only.

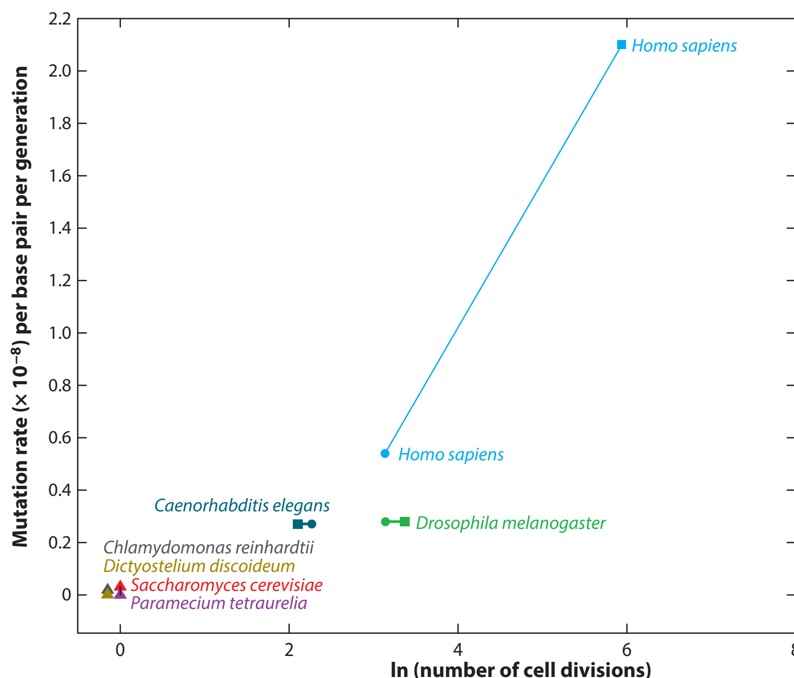

Figure 3

Direct estimates of the mutation rates in unicellular eukaryotes (*triangles*) and in female (*circles*) and male (*squares*) multicellular eukaryotes based on pedigree studies or mutation-accumulation experiments plotted against the log of the number of cell divisions from zygote to gamete. The number of cell divisions in *C. elegans* males and hermaphrodite females is from Drake et al. (31) and Kimble & Ward (74). The sex-averaged mutation rate in *C. elegans* was reported by Denver et al. (29a). The number of cell divisions in humans was taken to be $38 + 23 \times (30 - 13) = 429$, based on a mean male age of breeding of 30 years and on information from Drost & Lee (33). The number of cell divisions from zygote to egg is assumed to be independent of the female age in humans (as in 28, 33). The mutation rate in human males and females is based on phased mutations from Kong et al. (80), for which the mean age of fathers at proband birth was 30.5. The number of cell divisions for *D. melanogaster* is based on information from Drost & Lee (33). Following Drost & Lee (33), we assumed that the first egg stem cell division in *Drosophila* occurs at 8.5 days of development; unlike them, we assumed that females and males have a generation time of 14 days in the laboratory. The sex-averaged mutation rate in *D. melanogaster* was reported by Keightley et al. (72). For the unicellular eukaryotes [*C. reinhardtii* (115a), *D. discoideum* (141a), *S. cerevisiae* (96a), and *P. tetraurelia* (153a)], points are distributed around 0 so that they do not lie on top of one another.

(**Figure 2**). Moreover, there is no discernible maternal age effect (71, 80, 103). In fact, Kong et al. (80) reported that almost all the variance in mutation rates across offspring is explained by paternal age at birth. These findings are consistent with additional mutations being introduced in males through the 23 yearly cell divisions after puberty (61).

If all mutations were replication driven and errors occurred at the same rate per cell division in the two sexes and throughout ontogenesis, then the male bias should reflect the ratio of male to female cell divisions, c, which is thought to be 9.6 if conception occurs at age 25 and 13.3 if it occurs at age 30 (**Figure 4**). Yet α estimates from pedigree studies are only fourfold lower at age 30 (including for non-CpGs alone; **Figure 4**). This discrepancy may partly reflect error in c: The number of cell divisions before puberty, for instance, is inferred from static cell counts in ovarian

REVIEWS IN ADVANCE
Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

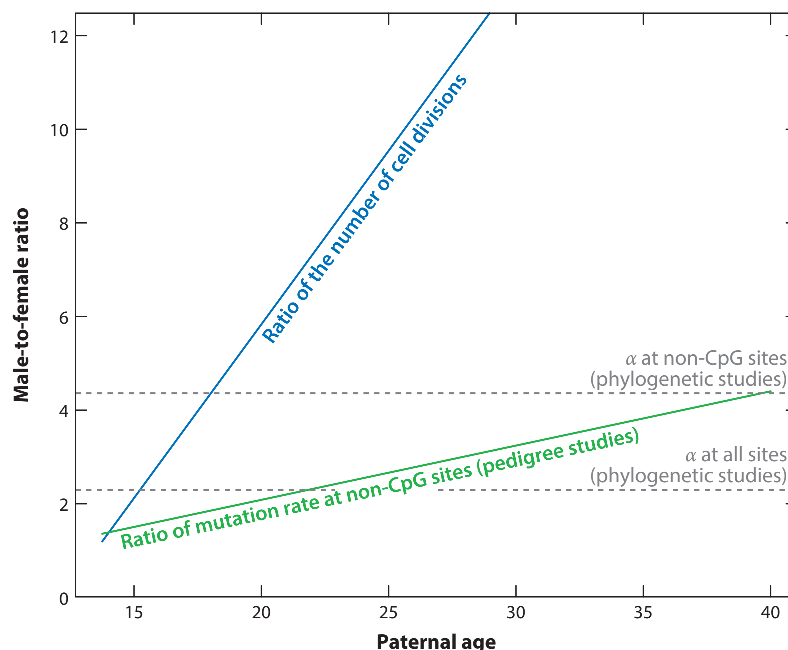

Figure 4

Male-to-female ratios of mutation rate and number of cell divisions as functions of paternal age. The blue line shows the ratio of the number of cell divisions in the male germline to that in the female germline, c, as a function of paternal age. We considered the total number of cell divisions to be 31 in females and 38 + (23/year) after puberty in males (33) and set puberty at age 13 (116). The green line is given by the slope estimated from the counts of phased non-CpG mutations in 15 trios, assuming no maternal age effect (19, 103). The intercept (i.e., the ratio of male to female mutations at puberty) was calculated based on the subset of non-CpG mutations that were assigned to the male and female germlines; we used the estimated slope for the paternal age effect for non-CpGs to subtract the expected number of paternal mutations that occurred after puberty. The dotted lines shows the phylogenetic estimate of the ratio of male to female mutations, α, for non-CpGs obtained by Xu et al. (170) based on the divergence in the human lineage since the split with orangutans, and the estimate of α for all sites obtained by Scally et al. (143) based on the divergence in the human lineage since the split with chimpanzees.

and testicular tissues under the assumption of symmetric cell divisions (159), and, similarly, the number of cell divisions in males after puberty depends on the model for spermatogenesis in humans (65, 130).

Another reason for α not to track c would be a maternal age effect on the number of cell divisions. Although there are no oogonial cell divisions after birth, all oocytes do not necessarily undergo the same number of divisions. Oocytes ovulated early in life may have undergone fewer mitotic divisions from the primordial germ cells than those ovulated later in life (64, 135), introducing the possibility of a maternal age effect on the number of point mutations. The difference in the number of cell divisions between early- and late-ovulated oocytes is unknown in humans, but it is expected to be small (135). In that regard, it is plausible that sequencing more families may yet reveal a maternal age effect on the number of maternal de novo mutations (46). This effect is expected to be subtle compared with the paternal one, however. An alternative explanation for the observation that α is much less than c is a higher error rate per cell division before puberty, i.e., if mutations accrue more slowly post-puberty in males (see discussion in sidebar Do We Expect a Generation Time Effect in Humans?).

REVIEWS IN ADVANCE
Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

Spontaneous mutations. In contrast to replication-driven errors, transitions that arise from spontaneous deamination and are largely independent of replication should accumulate proportionally to calendar time (assuming that methylation levels remain constant with age). Consistent with this expectation, the per-year substitution rate of transitions at CpGs is relatively constant compared with the rate at non-CpGs, across a range of species with different generation times (69, 73). In that regard, it would be interesting to assess whether divergence resulting from transversions at CpGs, whose source remains unknown, also shows a clocklike behavior.

If there are no sex differences in methylation, deamination rates, or repair efficacy, and assuming parents are similarly aged, one might further predict that α for spontaneous changes would be 1 at CpGs methylated in both sexes (8). Testing this prediction directly requires a sufficient number of de novo transitions at CpGs assigned to maternal and paternal germlines, and such data are not currently available. However, a small study of de novo mutation in chimpanzees found no difference between α at CpG sites and α at non-CpG sites (considering both transitions and transversions) (O. Venn, I. Turner, I. Mathieson, N. de Groot, Z. Iqbal, et al., manuscript in review), and phylogenetic studies also suggest that $\alpha > 1$ (155). An $\alpha > 1$ for transitions at CpGs could result from the fact that a subset of CpG mutations are replication driven if hydrolytic deamination of methylated CpGs in mammals is facilitated by strand separation [as observed for nonmethylated Cs in bacteria (47)]. Whether a substantial contribution of replication-driven transitions at CpGs is consistent with only a small generation time effect across mammals (69, 73) remains to be quantified. An alternative explanation could be a lower rate of spontaneous deamination (or better repair) in females than in males, so that mutations accumulate proportionally to absolute time in both sexes but less rapidly in females. This explanation would suggest low power to detect a maternal age effect for CpGs.

In summary, considering phylogenetic and pedigree studies together reveals considerable uncertainty about the sex-averaged mutation rate and the relative contributions of mutations from males compared with those from females. What is clear is that the value of α is much less than c, the ratio of male to female cell divisions. Together with the absence of a significant maternal age effect, this suggests one of two explanations. The first is that many mutations (at both CpG and non-CpG sites) are spontaneous in origin in both sexes (3, 4, 68) and that sperm are more sensitive to damage compared with eggs (5). This explanation implies a maternal age effect on mutation, albeit one that is too subtle to be detected with existing data. The second possibility is that the mutation rate per cell division decreases after puberty in males (see sidebar Do We Expect a Generation Time Effect in Humans?).

Other Sources of Interindividual Variation in the Mutation Rate

DNA damage, including lesions, adducts, and conformational changes, can result from exposure to various extrinsic agents (e.g., mutagens, UV and other radiation, and oxidative stress), which may vary among individuals. For instance, 1 gray of radiation exposure, equivalent to $\sim 3\times$ the average radiation exposure among atomic bomb survivors and on par with the dose received by some cancer patients, induces 0.27 mutational events in the coding regions of germlines (including substitutions, chromosomal aberrations, deletions, and translocations) (114). In addition to extrinsically induced differences, individuals may vary intrinsically in their propensity to acquire germline mutations, notably due to mutations in DNA mismatch repair genes implicated in hereditary cancer syndromes (2, 117, 165). Mismatch repair deficiencies are known to substantially elevate genome-wide per-generation mutation rates in yeast (82), but their effect on the germline mutation rate in humans remains to be assessed.

At CpG sites specifically, factors affecting individual variation in germline methylation level may induce variation in mutation rates (32). Approximately 70% of CpGs are methylated in

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

DO WE EXPECT A GENERATION TIME EFFECT IN HUMANS?

Thomas & Hahn (157) argued that we should not expect the mutation rate to depend on generation time in humans. For there to be no generation time effect, the number of mutations as a function of paternal age must be well described by a single line through the origin, when it is in fact piecewise linear. In other words, the slope must be the same before and after puberty, i.e.,

$$(\mu_0^{\mathrm{m}} c_0^{\mathrm{m}} + \mu_0^{\mathrm{f}} c_0^{\mathrm{f}})B/P = \mu_2^{\mathrm{m}} c_2 \tag{3}$$

If the mutation rate per cell division (μ^{m}) is constant throughout male development, as assumed by Thomas & Hahn (157), then this condition reduces to $\mu_0^{\mathrm{f}}/\mu^{\mathrm{m}} = (c_2 P - c_0^{\mathrm{m}} B)/c_0^{\mathrm{f}} B$. In humans, $B = 0.75$, and other plausible values are $P = 13.75$, $c_2 = 23$, $c_0^{\mathrm{m}} B = 34$, and $c_0^{\mathrm{f}} B = 31$, so the mutation rate per cell division in females would have to be precisely 9.10 times that in males from conception to birth, which seems even less likely when one considers that approximately half the cell divisions occur before sexual differentiation (33). Thus, all else being equal, if the mutation rate per cell division is the same throughout male development, then changes in the generation time, G, should affect the yearly mutation rate.

Assuming instead a similar mutation rate per cell division, μ_0, in males and females before birth, Equation 3 could only hold if $\mu_0/\mu_2 = c_2 P/(c_0^{\mathrm{m}} + c_0^{\mathrm{f}})B$, i.e., if the per-cell division mutation rate before and after puberty happens to take the precise value that satisfies the equation, which would seem highly coincidental. Thus, changes in the generation time seem highly likely to affect the yearly mutation rate in humans.

human sperm (106). No human oocyte methylomes are currently available, but in mice oocytes, CpG methylation levels are less than half those of sperm CpGs (77, 150). Sex-specific differences in germline methylation status would predict that males have a higher age-specific CpG mutation rate than females (which is consistent with the tentative evidence that α exceeds 1 at CpG sites). Environmental sources of variation such as diet are also known to affect methylation patterns (81, 149) and hence could in principle impact mutation rates in gametes.

THE EVOLUTION OF THE MUTATION RATE

Theoretical Basis

Intuitively, because most mutations are either neutral or deleterious, natural selection should act to decrease the mutation rate until it is counterbalanced by the cost of high-fidelity replication or limited by genetic drift. This net advantage of such mutation rate modifiers may differ across environments and ecological niches. Whether selection on such modifiers could be effective in humans is unclear.

In diploid species with recombination, the advantage of a mutation modifier is $D\bar{s}$, where D is the difference in the number of mutations per generation between diploid individuals with and without the M modifier and $\bar{s}$ is the average selection coefficient against a mutation in heterozygotes (31, 75, 78). To reach fixation in the population, the effects of the modifier must overcome the limits imposed by genetic drift—i.e., $D\bar{s}$ must exceed $1/(2N_e)$ (where N_e is the effective population size). Assuming a plausible value for $\bar{s}$ (e.g., 10^{-5} across all sites in the genome) and an N_e on the order of 10,000, a mutation modifier could be selected for so long as $D > 5$, which seems possible, depending on the costs of more effective repair.

This reasoning suggests that, all else being equal, species with smaller effective population sizes would experience a higher per-generation mutation rate (94, 95). Assuming that rates equilibrate

REVIEWS IN ADVANCE
Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

rapidly, then the low N_e of humans could have resulted in higher mutation rates compared with those of other vertebrates (85).

Other selection pressures may affect the mutation rate indirectly. Notably, selection on life history traits such as the mating system or age of reproduction could shape the mutation rate by leading to changes in the number of cell divisions per year. Among vertebrates, two important factors may be the generation time (141b, 170) and degree of sperm competition (133).

Under the generation time hypothesis, species with shorter generation times have a higher mutation rate per unit time (and hence a higher neutral substitution rate) because they undergo more germline replications in that fixed interval (89, 122) (for the specific conditions under which this holds, see sidebar The Effect of Generation Time on the Mutation Rate Per Year). As the average age of reproduction lengthens, replication-driven mutations will also become more male biased (22). This bias will further increase if puberty occurs early relative to the average age of male reproduction. Any factors that modulate the onset of puberty or mean age of reproduction will therefore affect the sex-averaged species mutation rate, especially in species with long reproductive life spans.

Selection on mating system also affects the number of male germline divisions. The intensity of sperm competition, i.e., the male–male contest over fertilization (123), shapes sperm production rates. Species with high levels of sperm competition typically have larger relative testis sizes to accommodate increased sperm production (134), and hence may also have higher male mutation rates per unit time after puberty. Although we might expect a positive correlation between the rate of sperm production and mutation rates, strong sexual selection against male germline mutations may at least partially counteract the increase in the mutation rate resulting from the increased number of cell divisions (166).

Selection may also act directly on the somatic mutation rate per cell division and, in so doing, affect the germline mutation rate through their shared repair pathways (94, 157). Additionally, germline mutation rates may be affected by factors such as metabolic rates if a higher metabolism causes more oxidative damage (69, 99). Selection against replication errors or oxidative damage in somatic tissues may be especially important for the viability and successful reproduction of long-lived, multicellular organisms (reviewed in 16). Disentangling the relative importance of different selection pressures on the mutation rate has proven difficult. Below, we review evidence bearing on how mutation rates differ among primates and the relationship to the number of germline cell divisions.

Available Data in Primates

A recent survey of 32 mammalian genomes confirmed that, as posited by the generation time hypothesis, the average age of reproduction (as measured by age at sexual maturity and life span) is negatively correlated with autosomal substitution rates (141b). For instance, rodents reach sexual maturity in months (89) and have a putatively neutral substitution rate per year that is 2–3 times higher than the rate in primates (169).

Generation time effects on substitutions may also be apparent within primates (126, 172). The generation times of the Old World monkeys baboons and rhesus macaques (~11 years) are approximately half that of chimpanzees and one-third that of humans (49). Accordingly, although fossils suggest similar divergence times between baboons and rhesus macaques and between humans and chimpanzees, the Old World monkeys have a 30% longer branch length at non-CpG sites (but similar levels of divergence at CpG sites) (40, 41, 73). Within apes, chimpanzees may have a slightly faster molecular clock than humans (41).

These observations suggest that humans have undergone an evolutionary slowdown relative to other primates, possibly driven by the lengthening of their generation time (41, 50, 142, 172). It is

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

worth noting, however, that it is the time between puberty and mean breeding age rather than the generation time per se that should be most relevant (see sidebars), and this interval may actually be longer in chimpanzees, for example, than in humans (43, 83, 98, 116).

The number of spermatogenetic cell divisions can be increased not only by a longer generation time relative to puberty but also by an increased intensity of sperm competition. Presgraves & Yi (133) suggested that relative testis mass, a proxy for the degree of sperm competition (57), is positively associated with an increase in the ratio of male to female mutations. Gorillas are polygynous (single male to many females) and have a small relative testis size (30). Humans are largely monogamous (single male per female) and have an intermediate relative testis size. By contrast, chimpanzees are promiscuous (multiple males per female) and have a large relative testis size, reflecting the fact that males produce many sperm per ejaculate. Chimpanzees produce sperm at a faster rate [every 14 days (151)] than do humans [every 16 days (61)] (the sperm production rates of gorillas are not known). These sperm production differences may account for a slightly stronger paternal age effect in chimpanzees (3 mutations per year; O. Venn, I. Turner, I. Mathieson, N. de Groot, Z. Iqbal, et al., manuscript in review). It is unclear, however, whether testis size per se helps to explain variation in α across primates (170) and mammals (141b), as it may be a poor measure of sperm competition (11, 107).

Other aspects of spermatogenesis may also evolve rapidly. In mice, one spermatogonial stem cell is thought to give rise to up to 4,096 haploid spermatids, whereas in rhesus macaques, it gives rise to up to 128 haploid spermatids, and in humans, it gives rise to only up to 16 haploid spermatids (38). The differences among the three species lie in the number of mitotic cell divisions prior to meiosis and may also contribute to the variation in mutation rates across species.

These considerations raise the possibility that changes in life history over the course of hominoid evolution have altered the number of spermatogenetic cell divisions per year. Any combination of a lengthening of the generation time, a delayed onset of puberty, or a postpubertal decrease in the number of mitotic cell divisions in males could have led to a decrease in the yearly mutation rate and help to explain why (contemporary) pedigree-based rates are lower than phylogenetic ones.

The possibility that mutation rates in hominoids are rapidly evolving introduces uncertainty about the dating of demographic events inferred from genetic data (e.g., 88) and the interpretation of divergence patterns on sex chromosomes and autosomes (124, 133, 161). In the short term, one way to circumvent these difficulties might be to focus on transitions at CpG sites, which constitute almost a fifth of all de novo mutations (80) and appear to be much less sensitive to the number of cell divisions per year (69). Longer term, in order to dissect the sources of interindividual variation in mutation rates and understand the factors that shape their evolution, there will be a continuing need to synthesize findings from biochemistry, evolutionary biology, and human genetics.

FUTURE ISSUES

1. To resolve uncertainty about the contemporary mutation rate and learn about sources of interindividual variation, we need additional pedigree studies that include more than one offspring and that are mindful of the number of sites reliably surveyed. Including a third generation will further allow one to distinguish somatic and germline mutations and to assign mutations to male and female germlines.
2. It will also be helpful to use new, complementary approaches to estimate the mutation rate. One approach is to compare the divergence on a lineage leading to securely dated hominoid fossils with the divergence to extant humans (48).

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

3. To learn about sources of spontaneous deamination, it will be important to study the methylomes of spermatogonial stem cells and arrested oocytes.
4. To learn how and why mutation rates evolve, it will be of interest to conduct pedigree-based studies of the mutation rate in other primate species, notably ones in which spermatogenesis is relatively well characterized (e.g., rhesus macaques).
5. It will also be of interest to obtain estimates of α across primates from context-dependent phylogenetic models, broken down by the likely mechanistic sources of mutation.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

Parental age data were obtained from the Simons Simplex Collection. We are grateful to A. Kong, R. Yuen, S. Sanders, C. Campbell, J. Michaelson, and A. Leotta for providing us with ancillary data used in **Figures 1**, **2**, and **4**, and to G. McVean, S. Sunyaev, and O. Venn for generously sharing their unpublished results. We also thank G. Amster, G. McVean, J. Pritchard, D. Reich, G. Sella, and especially Z. Gao and P. Moorjani for helpful discussions, as well as G. Amster, P. Green, P. Moorjani, G. Sella, and S. Sunyaev for helpful comments on an earlier version of the article. The work was conducted while M.P. was a Howard Hughes Medical Institute Early Career Scientist.

LITERATURE CITED

1. 1000 Genomes Proj. Consort. 2010. A map of human genome variation from population-scale sequencing. *Nature* 467:1061–73
2. Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, et al. 1999. Cancer risk in mutation carriers of DNA mismatch repair genes. *Int. J. Cancer* 81:214–18
3. Agarwal A, Aponte-Mellado A, Premkumar BJ, Shaman A, Gupta S. 2012. The effects of oxidative stress on female reproduction: a review. *Reprod. Biol. Endocrinol.* 10:49
4. Aitken RJ, De Iuliis GN. 2007. Origins and consequences of DNA damage in male germ cells. *Reprod. BioMed. Online* 14:727–33
5. Aitken RJ, De Iuliis GN, Gibb Z, Baker MA. 2012. The Simmet lecture: new horizons on an old landscape—oxidative stress, DNA damage and apoptosis in the male germ line. *Reprod. Domest. Anim.* 47(Suppl. 4):7–14
6. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, et al. 2013. Signatures of mutational processes in human cancer. *Nature* 500:415–21
7. Arndt PF, Burge CB, Hwa T. 2003. DNA sequence evolution with neighbor-dependent mutation. *J. Comput. Biol.* 10:313–22
8. Arnheim N, Calabrese P. 2009. Understanding what determines the frequency and pattern of human germline mutations. *Nat. Rev. Genet.* 10:478–88
9. Awadalla P, Gauthier J, Myers RA, Casals F, Hamdan FF, et al. 2010. Direct measure of the de novo mutation rate in autism and schizophrenia cohorts. *Am. J. Hum. Genet.* 87:316–24
10. Baer CF, Miyamoto MM, Denver DR. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat. Rev. Genet.* 8:619–31

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

11. Bartosch-Härlid A, Berlin S, Smith NGC, Møller AP, Ellegren H. 2003. Life history and the male mutation bias. *Evolution* 57:2398
12. Benzer S. 1961. On the topography of the genetic fine structure. *Proc. Natl. Acad. Sci. USA* 47:403–15
13. Bessman MJ, Muzyczka N, Goodman MF, Schnaar RL. 1974. Studies on the biochemical basis of spontaneous mutation. II. The incorporation of a base and its analogue into DNA by wild-type, mutator and antimutator DNA polymerases. *J. Mol. Biol.* 88:409–21
14. Bird AP. 1980. DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 8:1499–504
15. Bloom LB, Otto MR, Eritja R, Reha-Krantz LJ, Goodman MF, Beechem JM. 1994. Pre-steady-state kinetic analysis of sequence-dependent nucleotide excision by the 3′-exonuclease activity of bacteriophage T4 DNA polymerase. *Biochemistry* 33:7576–86
16. Bromham, L. 2011. The genome as a life-history character: why rate of molecular evolution varies between mammal species. *Philos. Trans. R. Soc. B* 366:2503–13
17. Brutlag D, Kornberg A. 1972. Enzymatic synthesis of deoxyribonucleic acid: XXXVI. A proofreading function for the $3' \rightarrow 5'$ exonuclease activity in deoxyribonucleic acid polymerases. *J. Biol. Chem.* 247:241–48
18. Bzymek M, Lovett ST. 2001. Instability of repetitive DNA sequences: the role of replication in multiple mechanisms. *Proc. Natl. Acad. Sci. USA* 98:8319–25
19. Campbell CD, Chong JX, Malig M, Ko A, Dumont BL, et al. 2012. Estimating the human mutation rate using autozygosity in a founder population. *Nat. Genet.* 44:1277–81
20. Campbell CD, Eichler EE. 2013. Properties and rates of germline mutations in humans. *Trends Genet.* 29:575–84
21. Capra JA, Hubisz MJ, Kostka D, Pollard KS, Siepel A. 2013. A model-based analysis of GC-biased gene conversion in the human and chimpanzee genomes. *PLoS Genet.* 9:e1003684
22. Chang BH, Shimmin LC, Shyue SK, Hewett-Emmett D, Li WH. 1994. Weak male-driven molecular evolution in rodents. *Proc. Natl. Acad. Sci. USA* 91:827–31
23. Chen F-C, Li W-H. 2001. Genomic divergences between humans and other hominoids and the effective population size of the common ancestor of humans and chimpanzees. *Am. J. Hum. Genet.* 68:444–56
24. Chimpanzee Seq. Anal. Consort. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87
25. Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, et al. 2011. Variation in genome-wide mutation rates within and between human families. *Nat. Genet.* 43:712–14
26. Cooper DN, Youssoufian H. 1988. The CpG dinucleotide and human genetic disease. *Hum. Genet.* 78:151–55
27. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W. 1978. Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 274:775–80
28. Crow JF. 2000. The origins, patterns and implications of human spontaneous mutation. *Nat. Rev. Genet.* 1:40–47
29. Crow JF. 2006. Age and sex effects on human mutation rates: an old problem with new complexities. *J. Radiat. Res.* 47(Suppl. B):B75–82

29a. Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, et al. 2009. A genome-wide view of *Caenorhabditis elegans* base-substitution mutation processes. *Proc. Natl. Acad. Sci. USA* 106:16310–14

30. Dixson AF, Anderson MJ. 2004. Sexual behavior, reproductive physiology and sperm competition in male mammals. *Physiol. Behav.* 83:361–71
31. Drake JW, Charlesworth B, Charlesworth D, Crow JF. 1998. Rates of spontaneous mutation. *Genetics* 148:1667–86
32. Driscoll DJ, Migeon BR. 1990. Sex difference in methylation of single-copy genes in human meiotic germ cells: implications for X chromosome inactivation, parental imprinting, and origin of CpG mutations. *Somat. Cell Mol. Genet.* 16:267–82
33. Drost JB, Lee WR. 1995. Biological basis of germline mutations: comparisons of spontaneous germline mutation rates among *Drosophila*, mouse, and human. *Environ. Mol. Mutagen.* 25(Suppl. S2):48–64
34. Duncan BK, Miller JH. 1980. Mutagenic deamination of cytosine residues in DNA. *Nature* 287:560–61

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

35. Duret L, Galtier N. 2009. Biased gene conversion and the evolution of mammalian genomic landscapes. *Annu. Rev. Genomics Hum. Genet.* 10:285–311
36. Ebersberger I, Metzler D, Schwarz C, Pääbo S. 2002. Genomewide comparison of DNA sequences between humans and chimpanzees. *Am. J. Hum. Genet.* 70:1490–97
37. Echols H, Goodman MF. 1991. Fidelity mechanisms in DNA replication. *Annu. Rev. Biochem.* 60:477–511
38. Ehmcke J. 2006. Spermatogonial stem cells: questions, models and perspectives. *Hum. Reprod. Update* 12:275–82
39. Ehrlich M, Wang RY. 1981. 5-Methylcytosine in eukaryotic DNA. *Science* 212:1350–57
40. Elango N, Lee J, Peng Z, Loh YHE, Yi SV. 2009. Evolutionary rate variation in Old World monkeys. *Biol. Lett.* 5:405–8
41. Elango N, Thomas JW, NISC Comp. Seq. Program, Yi SV. 2006. Variable molecular clocks in hominoids. *Proc. Natl. Acad. Sci. USA* 103:1370–75
42. Ellegren H. 2007. Characteristics, causes and evolutionary consequences of male-biased mutation. *Proc. Biol. Sci.* 274:1–10
43. Fenner JN. 2005. Cross-cultural estimation of the human generation interval for use in genetics-based population divergence studies. *Am. J. Phys. Anthropol.* 128:415–23
44. Fitch WM. 1967. Evidence suggesting a non-random character to nucleotide replacements in naturally occurring mutations. *J. Mol. Biol.* 26:499–507
45. Fousteri M, Mullenders LH. 2008. Transcription-coupled nucleotide excision repair in mammalian cells: molecular mechanisms and biological effects. *Cell Res.* 18:73–84
46. Fromer M, Pocklington AJ, Kavanagh DH, Williams HJ, Dwyer S, et al. 2014. De novo mutations in schizophrenia implicate synaptic networks. *Nature* 506:179–84
47. Fryxell KJ, Zuckerkandl E. 2000. Cytosine deamination plays a primary role in the evolution of mammalian isochores. *Mol. Biol. Evol.* 17:1371–83
48. Fu Q, Mittnik A, Johnson PL, Bos K, Lari M, et al. 2013. A revised timescale for human evolution based on ancient mitochondrial genomes. *Curr. Biol.* 23:553–59
49. Gage TB. 1998. The comparative demography of primates: with some comments on the evolution of life histories. *Annu. Rev. Anthropol.* 27:197–221
50. Goodman M. 1961. The role of immunologic differences in the phyletic development of human behavior. *Hum. Biol.* 33:131–62
51. Goodman MF, Creighton S, Bloom LB, Petruska J. 1993. Biochemical basis of DNA replication fidelity. *Crit. Rev. Biochem. Mol. Biol.* 28:83–126
52. Goodman MF, Fygenson KD. 1998. DNA polymerase fidelity: from genetics toward a biochemical understanding. *Genetics* 148:1475–82
53. Goriely A, Wilkie AO. 2012. Paternal age effect mutations and selfish spermatogonial selection: causes and consequences for human disease. *Am. J. Hum. Genet.* 90:175–200
54. Green P, Ewing B, Miller W, Thomas PJ, NISC Comp. Seq. Program, Green ED. 2003. Transcription-associated mutational asymmetry in mammalian evolution. *Nat. Genet.* 33:514–17
55. Haldane JBS. 1935. The rate of spontaneous mutation of a human gene. *J. Genet.* 31:317–26
56. Haldane JBS. 1947. The mutation rate of the gene for haemophilia and its segregation ratios in males and females. *Ann. Eugen.* 13:262–71
57. Harcourt AH, Harvey PH, Larson SG, Short RV. 1981. Testis weight, body weight and breeding system in primates. *Nature* 293:55–57
58. Hardison RC, Roskin KM, Yang S, Diekhans M, Kent WJ, et al. 2003. Covariation in frequencies of substitution, deletion, transposition, and recombination during eutherian evolution. *Genome Res.* 13:13–26
59. Harris K, Nielsen R. 2013. *Error-prone polymerase activity causes multinucleotide mutations in humans*. Unpublished manuscript, arXiv 1312.1395
60. Harwood J, Tachibana A, Meuth M. 1991. Multiple dispersed spontaneous mutations: a novel pathway of mutation in a malignant human cell line. *Mol. Cell. Biol.* 11:3163–70
61. Heller DA, Clermont Y. 1963. Spermatogenesis in man: an estimate of its duration. *Science* 140:184–86

Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

62. Hellmann I, Ebersberger I, Ptak SE, Pääbo S, Przeworski M. 2003. A neutral explanation for the correlation of diversity with recombination rates in humans. *Am. J. Hum. Genet.* 72:1527–35
63. Hellmann I, Mang Y, Gu Z, Li P, de la Vega FM, et al. 2008. Population genetic analysis of shotgun assemblies of genomic sequences from multiple individuals. *Genome Res.* 18:1020–29
64. Henderson SA, Edwards RG. 1968. Chiasma frequency and maternal age in mammals. *Nature* 218:22–28
65. Hermann BP, Sukhwani M, Hansel MC, Orwig KE. 2010. Spermatogonial stem cells in higher primates: Are there differences from those in rodents? *Reproduction* 139:479–93
66. Hess ST, Blake JD, Blake RD. 1994. Wide variations in neighbor-dependent substitution rates. *J. Mol. Biol.* 236:1022–33
67. Hodgkinson A, Eyre-Walker A. 2011. Variation in the mutation rate across mammalian genomes. *Nat. Rev. Genet.* 12:756–66
68. Huttley GA, Jakobsen IB, Wilson SR, Easteal S. 2000. How important is DNA replication for mutagenesis? *Mol. Biol. Evol.* 17:929–37
69. Hwang DG, Green P. 2004. Bayesian Markov chain Monte Carlo sequence analysis reveals varying neutral substitution patterns in mammalian evolution. *Proc. Natl. Acad. Sci. USA* 101:13994–4001
70. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, et al. 2012. De novo gene disruptions in children on the autistic spectrum. *Neuron* 74:285–99
71. Jiang YH, Yuen RK, Jin X, Wang M, Chen N, et al. 2013. Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* 93:249–63
72. Keightley PD, Ness RW, Halligan DL, Haddrill PR. 2014. Estimation of the spontaneous mutation rate per nucleotide site in a *Drosophila melanogaster* full-sib family. *Genetics* 196:313–20
73. Kim SH, Elango N, Warden C, Vigoda E, Yi SV. 2006. Heterogeneous genomic molecular clocks in primates. *PLoS Genet.* 2:e163
74. Kimble J, Ward S. 1998. Germ-line development and fertilization. In *The Nematode* Caenorhabditis elegans, ed. WB Wood, pp. 191–213. Woodbury, NY: Cold Spring Harb. Lab. Press
75. Kimura M. 1967. On the evolutionary adjustment of spontaneous mutation rates. *Genet. Res.* 9:23–34
76. Kimura M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge, UK: Cambridge Univ. Press
77. Kobayashi H, Sakurai T, Miura F, Imai M, Mochiduki K, et al. 2013. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* 23:616–27
78. Kondrashov AS. 1995. Modifiers of mutation-selection balance: general approach and the evolution of mutation rates. *Genet. Res.* 66:53–69
79. Kondrashov AS. 2003. Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. *Hum. Mutat.* 21:12–27
80. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488:471–75
81. Lambrot R, Xu C, Saint-Phar S, Chountalos G, Cohen T, et al. 2013. Low paternal dietary folate alters the mouse sperm epigenome and is associated with negative pregnancy outcomes. *Nat. Commun.* 4:1–13
82. Lang GI, Parsons L, Gammie AE. 2013. Mutation rates, spectra, and genome-wide distribution of spontaneous mutations in mismatch repair deficient yeast. *G3* 3:1453–65
83. Langergraber KE, Prufer K, Rowney C, Boesch C, Crockford C, et al. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc. Natl. Acad. Sci. USA* 109:15716–21
84. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499:214–18
85. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, et al. 2012. Revisiting an old riddle: What determines genetic diversity levels within species? *PLoS Biol.* 10:e1001388
86. Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, et al. 2013. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science* 339:1578–82
87. Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet.* 18:337–40

REVIEWS IN ADVANCE
Changes may still occur before final publication online and in print

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org
by University of Winnipeg on 08/28/14. For personal use only.

88. Li H, Durbin R. 2011. Inference of human population history from individual whole-genome sequences. *Nature* 475:493–96
89. Li WH, Ellsworth DL, Krushkal J, Chang BH, Hewett-Emmett D. 1996. Rates of nucleotide substitution in primates and rodents and the generation-time effect hypothesis. *Mol. Phylogenet. Evol.* 5:182–87
90. Li WH, Wu CI, Luo CC. 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. *J. Mol. Evol.* 21:58–71
91. Lindahl T. 1993. Instability and decay of the primary structure of DNA. *Nature* 362:709–15
92. Lindahl T, Nyberg B. 1974. Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry* 13:3405–10
93. Liu P, Carvalho CM, Hastings PJ, Lupski JR. 2012. Mechanisms for recurrent and complex human genomic rearrangements. *Curr. Opin. Genet. Dev.* 22:211–20
94. Lynch M. 2008. The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180:933–43
95. Lynch M. 2010a. Evolution of the mutation rate. *Trends Genet.* 26:345–52
96. Lynch M. 2010b. Rate, molecular spectrum, and consequences of human mutation. *Proc. Natl. Acad. Sci. USA* 107:961–68

96a. Lynch M, Sung W, Morris K, Coffey N, Landry CR, et al. 2008. A genome-wide view of the spectrum of spontaneous mutations in yeast. *Proc. Natl. Acad. Sci. USA* 105:9272–77

97. Makova KD, Li WH. 2002. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* 416:624–26
98. Marson J, Meuris S, Cooper RW, Jouannet P. 1991. Puberty in the male chimpanzee: progressive maturation of semen characteristics. *Biol. Reprod.* 44:448–55
99. Martin AP, Palumbi SR. 1993. Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. USA* 90:4087–91
100. McVean GT, Hurst LD. 1997. Evidence for a selectively favourable reduction in the mutation rate of the X chromosome. *Nature* 386:388–92
101. McVicker G, Gordon D, Davis C, Green P. 2009. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* 5:e1000471
102. McVicker G, Green P. 2010. Genomic signatures of germline gene expression. *Genome Res.* 20:1503–11
103. Michaelson JJ, Shi Y, Gujral M, Zheng H, Malhotra D, et al. 2012. Whole-genome sequencing in autism identifies hot spots for de novo germline mutation. *Cell* 151:1431–42
104. Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T. 1987. Male-driven molecular evolution: a model and nucleotide sequence analysis. *Cold Spring Harb. Symp. Quant. Biol.* 52:863–67
105. Mohrenweiser HW, Wilson DM III, Jones IM. 2003. Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes. *Mutat. Res.* 526:93–125
106. Molaro A, Hodges E, Fang F, Song Q, McCombie WR, et al. 2011. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* 146:1029–41
107. Møller AP, Cuervo JJ. 2003. Sexual selection, germline mutation rate and sperm competition. *BMC Evol. Biol.* 3:1–11
108. Moran S, Ren RX, Kool ET. 1997. A thymidine triphosphate shape analog lacking Watson-Crick pairing ability is replicated with high sequence selectivity. *Proc. Natl. Acad. Sci. USA* 94:10506–11
109. Muzyczka N, Poland RL, Bessman MJ. 1972. Studies on the biochemical basis of spontaneous mutation. I. A comparison of the deoxyribonucleic acid polymerases of mutator, antimutator, and wild type strains of bacteriophage T4. *J. Biol. Chem.* 247:7116–22
110. Nachman MW. 2004. Haldane and the first estimates of the human mutation rate. *J. Genet.* 83:231–33
111. Nachman MW, Crowell SL. 2000. Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304
112. Nagylaki T. 1983. Evolution of a finite population under gene conversion. *Proc. Natl. Acad. Sci. USA* 80:6278–81
113. Nakamura J, Swenberg JA. 1999. Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer Res.* 59:2522–26

Changes may still occur before final publication online and in print

114. Nakamura N, Suyama A, Noda A, Kodama Y. 2013. Radiation effects on human heredity. *Annu. Rev. Genet.* 47:33–50
115. Neale BM, Kou Y, Liu L, Ma'ayan A, Samocha KE, et al. 2012. Patterns and rates of exonic de novo mutations in autism spectrum disorders. *Nature* 485:242–45
115a. Ness RW, Morgan AD, Colegrave N, Keightley PD. 2012. Estimate of the spontaneous mutation rate in *Chlamydomonas reinhardtii*. *Genetics* 192:1447–54
116. Nielsen T, Skakkebaek NE, Richardson DW, Darling JAB, Hunter WM, et al. 1986. Onset of the release of spermatozoa (spermarche) in boys in relation to age, testicular growth, pubic hair, and height. *J. Clin. Endocrinol. Metab.* 62:532–35
117. Nik-Zainal S, Wedge DC, Alexandrov LB, Petljak M, Butler AP, et al. 2014. Association of a germline copy number polymorphism of *APOBEC3A* and *APOBEC3B* with burden of putative APOBEC-dependent mutations in breast cancer. *Nat. Genet.* 46:487–91
118. O'Roak BJ, Deriziotis P, Lee C, Vives L, Schwartz JJ, et al. 2011. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nat. Genet.* 43:585–89
119. O'Roak BJ, Vives L, Girirajan S, Karakoc E, Krumm N, et al. 2012. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. *Nature* 485:246–50
120. Odenthal-Hesse L, Berg IL, Veselis A, Jeffreys AJ, May CA. 2014. Transmission distortion affecting human noncrossover but not crossover recombination: a hidden source of meiotic drive. *PLoS Genet.* 10:e1004106
121. Ohno M, Sakumi K, Fukumura R, Furuichi M, Iwasaki Y, et al. 2014. 8-Oxoguanine causes spontaneous de novo germline mutations in mice. *Sci. Rep.* 4:4689
122. Ohta T. 1993. An examination of the generation-time effect on molecular evolution. *Proc. Natl. Acad. Sci. USA* 90:10676–80
123. Parker GA. 1970. Sperm competition and its evolutionary consequences in the insects. *Biol. Rev.* 45:525–67
124. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. 2006. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 441:1103–8
125. Penrose LS. 1955. Parental age and mutation. *Lancet* 269:312–13
126. Perry GH, Reeves D, Melsted P, Ratan A, Miller W, et al. 2012. A genome sequence resource for the aye-aye (*Daubentonia madagascariensis*), a nocturnal lemur from Madagascar. *Genome Biol. Evol.* 4:126–35
127. Petruska J, Goodman MF. 1985. Influence of neighboring bases on DNA polymerase insertion and proofreading fidelity. *J. Biol. Chem.* 260:7533–39
128. Pfeifer GP. 2006. Mutagenesis at methylated CpG sequences. *Curr. Top. Microbiol. Immunol.* 301:259–81
129. Pfeifer GP, You YH, Besaratinia A. 2005. Mutations induced by ultraviolet light. *Mutat. Res.* 571:19–31
130. Phillips BT, Gassei K, Orwig KE. 2010. Spermatogonial stem cell regulation and spermatogenesis. *Philos. Trans. R. Soc. B* 365:1663–78
131. Pleasance ED, Cheetham RK, Stephens PJ, McBride DJ, Humphray SJ, et al. 2010. A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463:191–96
132. Pleasance ED, Stephens PJ, O'Meara S, McBride DJ, Meynert A, et al. 2010. A small-cell lung cancer genome with complex signatures of tobacco exposure. *Nature* 463:184–90
133. Presgraves DC, Yi SV. 2009. Doubts about complex speciation between humans and chimpanzees. *Trends Ecol. Evol.* 24:533–40
134. Ramm SA, Stockley P. 2010. Sperm competition and sperm length influence the rate of mammalian spermatogenesis. *Biol. Lett.* 6:219–21
135. Reizel Y, Itzkovitz S, Adar R, Elbaz J, Jinich A, et al. 2012. Cell lineage analysis of the mammalian female germline. *PLoS Genet.* 8:e1002477
136. Risch N, Reich EW, Wishnick MM, McCarthy JG. 1987. Spontaneous mutation and parental age in humans. *Am. J. Hum. Genet.* 41:218–48
137. Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328:636–39
138. Roberts JD, Kunkel TA. 1996. *Fidelity of DNA Replication*. Woodbury, NY: Cold Spring Harb. Lab. Press

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org by University of Winnipeg on 08/28/14. For personal use only.

REVIEWS IN ADVANCE

Changes may still occur before final publication online and in print

139. Rubin AF, Green P. 2009. Mutation patterns in cancer genomes. *Proc. Natl. Acad. Sci. USA* 106:21766–70
140. Ruggeri B, DiRado M, Zhang SY, Bauer B, Goodrow T, Klein-Szanto AJ. 1993. Benzo[*a*]pyrene-induced murine skin tumors exhibit frequent and characteristic G to T mutations in the p53 gene. *Proc. Natl. Acad. Sci. USA* 90:1013–17
141. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, et al. 2012. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 485:237–41

141a. Saxer G, Havlak P, Fox SA, Quance MA, Gupta S, et al. 2012. Whole genome sequencing of mutation accumulation lines reveals a low mutation rate in the social amoeba *Dictyostelium discoideum*. *PLoS ONE* 7:e46759

141b. Sayres MAW, Venditti C, Pagel M, Makova KD. 2011. Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution* 65:2800–15

142. Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nat. Rev. Genet.* 13:745–53
143. Scally A, Dutheil JY, Hillier LW, Jordan GE, Goodhead I, et al. 2012. Insights into hominid evolution from the gorilla genome sequence. *Nature* 483:169–75
144. Schaibley VM, Zawistowski M, Wegmann D, Ehm MG, Nelson MR, et al. 2013. The influence of genomic context on mutation patterns in the human genome inferred from rare variants. *Genome Res.* 23:1974–84
145. Schmidt S, Gerasimova A, Kondrashov FA, Adzhubei IA, Kondrashov AS, Sunyaev S. 2008. Hypermutable non-synonymous sites are under stronger negative selection. *PLoS Genet.* 4:e1000281
146. Schmutte C, Yang AS, Beart RW, Jones PA. 1995. Base excision repair of U:G mismatches at a mutational hotspot in the *p53* gene is more efficient than base excision repair of T:G mismatches in extracts of human colon tumors. *Cancer Res.* 55:3742–46
147. Seidman MM, Bredberg A, Seetharam S, Kraemer KH. 1987. Multiple point mutations in a shuttle vector propagated in human cells: evidence for an error-prone DNA polymerase activity. *Proc. Natl. Acad. Sci. USA* 84:4944–48
148. Siepel A, Haussler D. 2004. Phylogenetic estimation of context-dependent substitution rates by maximum likelihood. *Mol. Biol. Evol.* 21:468–88
149. Sinclair KD, Allegrucci C, Singh R, Gardner DS, Sebastian S, et al. 2007. DNA methylation, insulin resistance, and blood pressure in offspring determined by maternal periconceptional B vitamin and methionine status. *Proc. Natl. Acad. Sci. USA* 104:19351–56
150. Smallwood SA, Kelsey G. 2012. De novo DNA methylation: a germ cell perspective. *Trends Genet.* 28:33–42
151. Smithwick EB, Young LG, Gould KG. 1996. Duration of spermatogenesis and relative frequency of each stage in the seminiferous epithelial cycle of the chimpanzee. *Tissue Cell* 28:357–66
152. Stamatoyannopoulos JA, Adzhubei I, Thurman RE, Kryukov GV, Mirkin SM, Sunyaev SR. 2009. Human mutation rate associated with DNA replication timing. *Nat. Genet.* 41:393–95
153. Stone JE, Lujan SA, Kunkel TA, Kunkel TA. 2012. DNA polymerase zeta generates clustered mutations during bypass of endogenous DNA lesions in *Saccharomyces cerevisiae*. *Environ. Mol. Mutagen.* 53:777–86

153a. Sung W, Tucker AE, Doak TG, Choi E, Thomas WK, Lynch M. 2012. Extraordinary genome stability in the ciliate *Paramecium tetraurelia*. *Proc. Natl. Acad. Sci. USA* 109:19339–44

154. Sunyaev S, Polak P, Francioli L, Kloosterman W, de Bakker PIW, Genome Neth. Consort. 2013. *Large-scale parent-child trio sequencing highlights factors influencing spontaneous human mutation*. Presented at Am. Soc. Hum. Genet. Meet., Boston, October 22–26
155. Taylor J, Tyekucheva S, Zody M, Chiaromonte F, Makova KD. 2006. Strong and weak male mutation bias at different sites in the primate genomes: insights from the human-chimpanzee comparison. *Mol. Biol. Evol.* 23:565–73
156. Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, et al. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337:64–69
157. Thomas GWC, Hahn MW. 2014. The human mutation rate is increasing, even as it slows. *Mol. Biol. Evol.* 31:253–57

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org by University of Winnipeg on 08/28/14. For personal use only.

Changes may still occur before final publication online and in print

158. Trautner TA, Swartz MN, Kornberg A. 1962. Enzymatic synthesis of deoxyribonucleic acid. X. Influence of bromouracil substitutions on replication. *Proc. Natl. Acad. Sci. USA* 48:449–55
159. Vogel F, Rathenberg R. 1975. Spontaneous mutation in man. *Adv. Hum. Genet.* 5:223–318
160. Wakeley J. 1996. The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends Ecol. Evol.* 11:158–62
161. Wakeley J. 2008. Complex speciation of humans and chimpanzees. *Nature* 452:E3–4
162. Wang J, Fan HC, Behr B, Quake SR. 2012. Genome-wide single-cell analysis of recombination activity and de novo mutation rates in human sperm. *Cell* 150:402–12
163. Watanabe Y, Fujiyama A, Ichiba Y, Hattori M, Yada T, et al. 2002. Chromosome-wide assessment of replication timing for human chromosomes 11q and 21q: disease-related genes in timing-switch regions. *Hum. Mol. Genet.* 11:13–21
164. Weinberg W. 1912. Zur vererbung des zwergwuchses. *Arch. Rassen-u. Ges. Biol.* 9:710–18
165. Wheeler JMD, Bodmer WF. 2000. DNA mismatch repair genes and colorectal cancer. *Gut* 47:148–53
166. Whitlock MC, Agrawal AF. 2009. Purging the genome with sexual selection: reducing mutation load through selection on males. *Evolution* 63:569–82
167. Deleted in proof
168. Woodruff RC, Thompson JNJ. 1992. Have premeiotic clusters of mutation been overlooked in evolutionary theory? *J. Evol. Biol.* 5:457–646
169. Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* 82:1741–45
170. Xu K, Oh S, Park T, Presgraves DC, Yi SV. 2012. Lineage-specific variation in slow- and fast-X evolution in primates. *Evolution* 66:1751–61
171. Xue Y, Wang Q, Long Q, Ng BL, Swerdlow H, et al. 2009. Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* 19:1453–57
172. Yi S, Ellsworth DL, Li WH. 2002. Slow molecular clocks in Old World monkeys, apes, and humans. *Mol. Biol. Evol.* 19:2191–98
173. Ying H, Epps J, Williams R, Huttley G. 2010. Evidence that localized variation in primate sequence divergence arises from an influence of nucleosome placement on DNA repair. *Mol. Biol. Evol.* 27:637–49
174. Yoon JH, Smith LE, Feng Z, Tang M, Lee CS, Pfeifer GP. 2001. Methylated CpG dinucleotides are the preferential targets for G-to-T transversion mutations induced by benzo[*a*]pyrene diol epoxide in mammalian cells: similarities with the p53 mutation spectrum in smoking-associated lung cancers. *Cancer Res.* 61:7110–17
175. Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, et al. 2013. De novo mutations in histone-modifying genes in congenital heart disease. *Nature* 498:220–23
176. Zenzes MT. 2000. Smoking and reproduction: gene damage to human gametes and embryos. *Hum. Reprod. Update* 6:122–31

Annu. Rev. Genom. Human Genet. 2014.15. Downloaded from www.annualreviews.org by University of Winnipeg on 08/28/14. For personal use only.

Changes may still occur before final publication online and in print