Somatic maintenance alters selection acting on mutation rate

Authors:

Andrii I. Rozhok[1] and James DeGregori[1,2,3,4]

[1]Department of Biochemistry and Molecular Genetics, [2]Integrated Department of Immunology, [3]Department of Pediatrics, [4]Department of Medicine, Section of Hematology, University of Colorado School of Medicine, Aurora, CO 80045

Corresponding Authors:

Andrii I. Rozhok: Andrii.Rozhok@ucdenver.edu

James DeGregori: James.DeGregori@ucdenver.edu

Acknowledgments

We would like to thank Mark Johnston, L. Alex Liggett and David Pollock of the University of Colorado, and Robert Gatenby and Andriy Marusyk of the Moffitt Cancer Center for critical review of the manuscript. These studies were supported by National Cancer Institute grant R01CA180175 to J.D.

Abstract

The evolution of multi-cellular animals has produced a conspicuous trend toward increased body size. This trend has introduced at least two novel problems: an elevated risk of somatic disorders, such as cancer, and declining evolvability due to reduced population size, lower reproduction rate and extended generation time. Low population size is widely recognized to explain the high mutation rates in animals by limiting the presumably universally negative selection acting on mutation rates. Here, we present evidence from stochastic modeling that the direction and strength of selection acting on mutation rates is highly dependent on the evolution of somatic maintenance, and thus longevity, which modulates the cost of somatic mutations. We propose a theoretical model for how evolvability and germline mutation rates can be under positive selection in sexually reproducing organisms by their co-selection with adaptive alleles that overcomes gene segregation produced by genetic recombination. We argue that this mechanism may have been critical in facilitating animal evolution.

Keywords: somatic maintenance, longevity, body size, mutation rate, selection

Introduction

Increasing body size has been one of the major trends in animal evolution across many taxa, as formulated in Cope's rule (Heim et al., 2015, Baker et al., 2015). The evolution of larger bodies introduces some fundamentally new evolutionary challenges. The carrying capacity of ecosystems limits biomass per group/species, so larger body size leads to reduced population size. Furthermore, large animals generally demonstrate lower reproduction rates and longer generation times. In aggregate, such changes weaken selection that can act on a population and thus negatively affect evolvability. This general reduction in evolvability should, however, be at least partially alleviated by diversity facilitated by sexual reproduction.

The mutation rate (MR) is another critical evolvability parameter. It is believed that selection generally acts to lower MR (Kimura, 1967, Baer et al., 2007, Dawson, 1999), and the significantly higher MRs observed in animals compared to unicellular organisms have been argued to result from the reduced power of selection imposed by small population sizes (Lynch, 2010, Lynch, 2011, Lynch et al., 2016). Germline (gMR) and somatic (sMR) mutation rates are linked, as they employ the same basic DNA replication and repair machinery (Pothof et al., 2003, Marcon & Moens, 2005, Galetzka et al., 2007). While elevated gMR improves evolvability, the ensuing higher sMR should elevate the risk of somatic disorders, such as cancer (Hanahan & Weinberg, 2000). For cancer, increasing body size is expected to increase the frequency of oncogenic mutations by increasing the number of target cells (Caulin & Maley, 2011). Somatic mutations also contribute to aging and a variety of aging-related diseases (Lopez-Otin et al., 2013). The increased cost of sMR should thus exert negative selective pressure on gMR in larger animals.

Recent evidence demonstrates that the sMR in some animal tissues can be significantly higher than the rate inferred from observed mutations, because somatic purifying selection is very effective in eliminating damaged somatic cells (Pfau et al., 2016). Many mechanisms, such as various tumor suppressor gene functions (including DNA damage induced apoptosis) (Sherr, 2004), autophagy (Glick et al., 2010), purifying somatic selection (Pfau et al., 2016, Rozhok & DeGregori, 2016), and immune surveillance (Swann & Smyth, 2007), should buffer the costs of

somatic mutation and in aggregate promote lifespan extension by maintaining tissue integrity. We will collectively call these mechanisms – the *somatic maintenance program* (SMP).

We present theoretical evidence from Monte Carlo modeling indicating that somatic maintenance not only improves individuals' survival for large animals by reducing sMR costs, but should have played a crucial role in animal evolution by substantially modifying selection acting on gMR. We show that positive selection for increased body size promotes positive selection for extended longevity by improving SMP. Our results also indicate that positive selection on traits that do not impact somatic risks also promotes selection for an improved SMP. In both cases, positive selection on gMR was observed because of the reduced sMR cost, which dramatically improved evolvability of the simulated population. While high MR is always a disadvantageous trait on its own, we propose a model for how MR contributes to individual net fitness and how small population size promotes selection for higher evolvability by elevating gMR.

Methods

Software. The model was created and all simulations were run in the Matlab environment (MathWorks Inc, MA) version R2014a.

Model algorithm. The model is a stochastic Monte Carlo type model (the exact algorithm can be found in **Supplements: Section 1a**) that runs a total of 1,005,000 updates ("time" in arbitrary units, AU) unless otherwise stated, which represents ~1000 generations of the simulated animal population. The simulation starts with building an initial population of 10,000 individuals. Each individual has a number of simulated traits: 1) ID, which is 1 (monogenotypic population) or 1 and 2 (in experiments with competition between two genotypes in a mixed population to indicate genotypes); 2) current age, which increments by 1 at each simulation update; 3) inherited body mass, which is inherited with variation by an individual and will be reached by adulthood (at age ~1000) and equals 5000 AU in the initial population; 4) current body mass, which changes during individual growth, following a growth curve, and plateaus at the inherited body mass in adults; 5) inherited birth mass, which in individuals of the initial population is 300 AU; 6) inherited

mutation rate of 10^{-9} AU (explained below); 7) inherited reproduction rate, which is the period with variation between successive reproductions in adult individuals and equals ~600 in the initial population; 8) inherited litter size (initially 1), which is the number of progeny produced per individual per reproduction; 9) inherited parameter of somatic maintenance, which determines the strength of the somatic maintenance program as further explained below; 10) age of first reproduction, which dictates that an individual begins reproducing when its current body mass reaches 0.9693 of its inherited adult body mass (the number is derived so that in the initial population maturity is reached at age ~1000 based on the growth curve).

Each inherited trait varies in progeny relative to parental. This variation was produced by multiplying the inherited mutation rate by the parameter of inherited variance (*inhvar* = 250,000,000) and the product was used as the standard deviation (STD) of the normally distributed variation in inheritance. This transformation was not necessary, as the *inhvar* parameter is constant throughout simulation and it simply determines the magnitude of the mutation rate's effects in germline, which is imaginary and in the initial population simply produces 0.000000001 x 25,000,000 = 0.025 that serves as the STD parameter for the normal distribution from which inheritance variation is drawn. However, we kept this two-parametric model for inheritance because mutation rate is also separately used in the equation of the somatic maintenance program (as will be explained later).

Each newborn individual grows, reaches maturity, then reproduces over the rest of its lifetime and eventually dies. The model is asynchronous, so that at every time-point of the simulation the population contains individuals of various ages whose lifecycles develop independently. The model operates with single-parent reproduction model so that each individual descends from one parent. In this regard, technically it is tempting to view it as a model of an asexual population. However, at a higher level of abstraction the fundamental difference between sexual and asexual populations (aside from the issue of purging deleterious mutations) is the amount of variation produced per the same size population per generation. Variance of inheritance in our model (as shown above) is obviously too high to be assumed as being generated by mutations accumulating along a clonal lineage and equals 10% of a trait's value per generation within 95 percentile. As the modeled traits are assumed to be *multigenic* and have a continuous phenotypic range in the population, we did not need to simulate the processes of allelic

segregation by recombination in order to reconstruct a sexual population. As such, the model only operates with the net ultimate change of a trait over generations. At this level of abstraction, the effective difference between a sexual and asexual population is reduced to the amount of variation in phenotypically manifested inheritance per population size per generation. We account for population size in this definition by inferring that this variance per se will not depend on population size, but larger populations will have higher chance of generating extreme phenotypes, e.g. those beyond 95 percentile on a per generation basis.

And finally, three factors of mortality were modelled in the simulations. First, at every timepoint of the simulation, an individual could die of somatic causes with a certain probability. This probability is small at the beginning of life (but still can be caused by some imaginary inherited genetic defects) and increases exponentially with age based on the paradigm of the aging curve, which is primarily determined by an individual's inherited somatic maintenance program (SMP). In humans, the aging curve also depends on lifestyle, however we assume in this model that in a wild animal population lifestyle distribution is sufficiently uniform to be neglected. More detailed description of the somatic maintenance paradigm we applied will be explained further below. Secondly, the simulated animals had a chance of dying of external hazards, such as predators. We applied the Lotka-Volterra model of predator-prey interactions (Lotka, 1925, Volterra, 1926) to implement the dynamics of predator pressure (effectively the chance of dying of an external hazard cause per timeunit). Here we should mention that smaller individuals and juveniles had higher chances of dying of external hazards, which effectively created positive selection for body size and also reflected the typical high mortality rates among juveniles observed in natural populations. And lastly, individuals could die of intra-specific competition. We implemented such competition by setting the upper limit of population's total biomass, which in nature is imposed by the ecosystem's carrying capacity. Therefore, in the simulated population biomass produced over the biomass limit caused additional mortality so that stochastically population total biomass never exceeded the limit. Larger individuals also had lower probability of dying of intra-specific competition, based on the assumption that competition for resources and mates (the failure to reproduce is effectively an evolutionary death) will typically favor larger individuals and this should have been one of the forces that has been driving the macroscopic animal evolutionary trend towards increasing body size. The advantage of size in this mortality model also created

additional positive selective pressure for body size. The total age-dependent mortality of all causes in our model did approximate a typical wild animal mortality curve (**Supplements: Section 3**).

The somatic maintenance program paradigm. In order to replicate natural mortality caused by physiological aging, such as cancer, decreased immune defense and lower ability to avoid predators or to succeed in intra-specific competition, we made use of the aging curve, or somatic maintenance, concept. Modern humans (in developed nations) and captive animal mortality curves (**Fig. 1B** for human) differ from wild animal mortality curves in very high early life survival with most mortality significantly delayed into advanced ages (Hochberg & Noble, 2017, Madsen et al., 2017). This difference is caused by many reasons, such as much lower mortality caused by external hazards and better nutrition and general healthcare. It therefore can be assumed that the human and captive animal mortality curves are close representations of the physiological aging curve. As longevity depends on multiple mechanisms of maintaining the soma, we can also call this curve *the somatic maintenance curve*. In order to reconstruct this curve, we assumed that somatic maintenance depends on the interaction of two opposing forces: 1) the accumulation of genetic and structural damage in the soma that promotes aging and 2) the somatic maintenance program consisting of a number of mechanisms that prevent or buffer the effects of genetic and structural damage. The exact mathematical relationship between these two forces and age is not known, however an example of cancer development can be used as a proxy to explain the equation we derived for it. Oncogenic mutations (including oncogenic epigenetic changes) are the ultimate necessary condition for cancer to develop. The frequency of oncogenic mutations linearly depends on mutation rate on a per cell division basis. Therefore, we assume that linear changes in mutation rate will have linear effects on the odds of the occurrence of oncogenic mutations. An oncogenic mutation provides the initiated cells with a linear change in their fitness relative to normal cells. However, over time an advantageous clone with a constant linear fitness advantage will proliferate exponentially. Therefore, we can already assume that mutation rate should have a linear effect on the cancer curve, while time/age adds an exponential component revealed in an exponential growth of a tumor. We can reasonably assume further that a strong SMP will efficiently suppress such a clone, slowing or even preventing its growth. A weaker SMP will allow the clone to proliferate faster. Therefore, SMP

strength can modulate the effects of mutations and time on cancer risk. The exact relationship between SMP strength and physiological risk factors is not known. However, we know that their interaction leads to a net exponent in physiological decline and disease risk. We therefore reconstructed the human aging curve by maintaining the general principal relationship between these factors as shown in **Eq. 1**. As seen from the equation, mutation rate is a linear contributor to aging. Age itself contributes exponentially, and the somatic maintenance composite parameter *Som* is, in turn, in power relationship to age. The cumulative distribution function of D_A (**Eq. 1**) produces *D(A)* – the probability of dying of somatic/physiological causes by age *A* and yields a shape close to the human mortality curve (**Fig. 1A,B**). We cannot claim that these three factors are in the exact relationship predicted by **Eq. 1**, as it is unknown. As seen in **Fig. 1A**, changes in the *Som* parameter have substantially greater effects on the resulting mortality curve than mutation rate, with mutation rate still having a sizeable effect as well. Yet claims are still made(Kennedy et al., 2012) that mutation rate is a larger factor in aging than we assume in this model. Validation of our assumption in general comes from the body of solid evidence that up to 50% of mutations in humans accumulate during body growth by the age 18-20 (Finette et al., 1994, Giese et al., 2002, Horvath, 2013). If mutation accumulation had a significant effect on aging on its own, we should age rapidly until age 18-20 (half-way) and then the rate of aging should decelerate. However, in reality the opposite happens, indicating that the combined strength of the SMP has an overpowering effect in modulating the effects of genetic damage on aging. As a result, we reason that **Eq. 1** might reasonably approximate the natural relationships of these three factors. Therefore, based on an individual's aging curve we calculated the D_A parameter at each simulation time-point (using the individual's mutation rate, age and *Som* parameter) and applied it in a binomial trial as the probability of that individual's dying of somatic/physiological causes in an age-dependent manner. As further explained in **Supplements: Section 4**, the exact relationship between the *Som* parameters and each of the other two (mutation rate and age) has no effect on the model, as the model represents SMP and its variation by using area under the mortality curve, therefore the sole purpose of **Eq. 1** in the model is to generate an age-dependent curve of physiological mortality whose cumulative function (probability of dying by a certain age) resembles in shape the human mortality/aging curve (see **Supplements: Section 4** for detailed explanation and illustration).

Model variations. A number of model variations used in simulation experiments are employed. *Fixed trait values* involved simply fixing the initial trait value without inherited variation throughout the entire simulation. *Dislinking of somatic and germline mutation rate* was done by making the value *M* in **Eq. 1** independent of an individual's mutation rate, which resulted in somatic costs independent of transgenerational variation of mutation rate (effectively from germline mutation rate). *Selection for a trait that did not affect somatic risks* was achieved by transforming the "body mass" trait's effects by removing the trait from calculations of the risk of death by somatic causes (unlike body size, it did not influence the risk), then removing the population biomass limit and setting maximum population size (unlike body mass, other traits do not directly affect population numbers) and fixing the growth rate curve so that it reached the initial body mass of 5,000 AU (the current body mass parameter in the model; the inherited body mass variation did not exist and the inherited body mass parameter was replaced with the somatic risk unrelated trait). These manipulations made the selected trait a proxy for a trait unrelated to somatic risks (e.g. hair color). *Competitive assays* included individuals with different ID parameters, such as 1 and 2 to indicate different "genotypes"; traits of the "genotypes" then were tracked and stored separately.

Data processing. Processing of primary data included removal of outliers (see **Supplements: Section 5**). Occasionally the simulations generated "NaN" (not a number) values in individual parameters, which were rare but quickly propagated if left in the population. We immediately deleted individuals from the population if "NaN" values appeared in any of their parameters. Based on the rarity of such events, we can assume that they had the effect of rare early lethal mutations and affected the population at random. Thus we assume these did not affect the principal results.

Statistics and data presentation. Most simulation experiments were made with 25 repeats. Due to heavy skews in sample distributions (inferred by D'Agostino-Pearson test for normality of a distribution), all figure panels represent medians (thick lines) and 95 percentiles on each tail (color-shaded areas). Statistical differences between experimental conditions were calculated as follows. We first calculated the sum of all values in each run throughout the entire evolution

of a trait (typically 1,005,000 time points). In this way, given the small increment over a long time the sum essentially approximated the area under the curve of a trait's evolution. These sums (usually 25 repeats in one experiment/sample) were then compared by applying the Matlab implementation of the Wilcoxon rank sum test, which is considered equivalent to the Mann-Whitney U-test. P-values <= 0.05 were considered as indicating significant difference.

Results

We built a stochastic model of evolution in animal populations, incorporating reproduction and survival, whereby each individual's traits are inherited with variance proportional to gMR (for code, see **Supplements: Section 1a**). Traits are assumed to be polygenic and exhibit phenotypic variation in the population. The evolution of body size, somatic maintenance and germline mutation rate was then tracked under various regimens of selection. The model reasonably approximates a sexually reproducing population as explained in **Methods: Model algorithm**.

The model incorporates three major factors of mortality, including aging. Human life tables indicate that aging proceeds exponentially, whereby mortality and diseases accelerate at advanced ages (e.g. https://www.ssa.gov, https://seer.cancer.gov). The combined action of SMP mechanisms provides for an extended early period of high body fitness with little to no decline. We generalized this complex program in a curve that describes modeled animal mortality of physiological causes schematically shown in **Fig. 1A** and based on the following equation:

$$D_A = M \times e^{A^{Som}} \qquad (1)$$

where D_A is the probability of dying of physiological causes at age A, M is mutation rate, and Som is a composite parameter that determines SMP efficiency. The cumulative distribution function of D_A, or the probability of dying of physiological causes by age A, resembles human mortality (**Fig. 1B**). The equation should thus provide a robust model for aging-related mortality, reflecting the extended period of high fitness and the late-life accelerating mortality. **Fig. 1A** also demonstrates the relative effects of MR, which is a linear contributor, and the *Som* parameter,

which stands for the total damage buffering capacity of the SMP (for details and theory see **Methods: The somatic maintenance program paradigm**). It is important to keep in mind that the *M* parameter (mutation rate) in **Eq. 1** is responsible for the somatic costs of MR (higher MR in **Fig. 1A** accelerates aging-related mortality).

In our simulations, positive selection for body size (**Fig. 1C, green**) led to a concurrent selection for elevated gMR (**Fig. 1D, green**) and improved SMP (**Fig. 1E, green**). Artificially blocking SMP evolution by fixing SMP at the initial value (**Fig. 1E, blue**) significantly slowed the evolution of body size (**Fig. 1C, blue; $p << 0.001$**) and triggered negative selection on gMR (**Fig. 1D, blue**). We implemented the ecosystem carrying capacity by setting a maximum biomass for the population; therefore, increasing body size led to a corresponding decline in population numbers, amplifying the power of drift (**Fig. 1F,G**). When SMP was allowed to evolve, however, the population entered a "drift zone" when its size decreased to ~4,000 individuals, which shortly thereafter was overcome by selection for even larger body size, visible also by a continuing decline in population numbers (**Fig. 1F**). When we artificially blocked SMP, however, the drift zone was more profound. It occurred earlier at the population size of ~6,000-7,000 individuals, and the population was not able to escape from it (for ~1,000 generations) and restore its initial rates of evolution (**Fig. 1G**), indicating an important role of SMP evolution in maintaining evolvability. We further generated a population with two simulated genotypes – Genotype A that could evolve SMP (10% of the population) and Genotype B with SMP fixed at the initial value (90%). We set a maximum population size and removed the maximum biomass limit to rule out body mass effects on population size and selection, and tracked Genotype A and Genotype B frequencies under positive selection for body size (for code see **Supplements: Section 1b**). Despite the initial abundance, Genotype B (with fixed SMP) lost the competition in less than 200 generations, reflecting a direct competitive advantage of the capacity to evolve enhanced SMP (**Fig. 1H**). Hereafter, we will call the setting with positive selection for body size and freely evolving SMP and gMR the *standard condition* (usually shown in green, unless otherwise indicated) used in comparisons with other selection regimens.

In the absence of positive selection for increased body mass (**Fig. 2A, blue**), both gMR (**Fig. 2B, blue**) and SMP (**Fig. 2C, blue**) demonstrate early positive selection, which appeared to have been caused by rapid evolution of reproductive parameters (see **Supplement: Section 2**). Overall, gMR demonstrates a significant general decrease (non-overlapping confidence intervals (CIs) at the beginning relative to the end of the simulation), and SMP undergoes a significantly smaller improvement compared to the standard condition (green; $p \ll 0.001$). Blocking the evolution of body mass (**Fig. 2D, blue**) and SMP (**Fig. 2F, blue**) expectedly led to strong selection for lower gMR (**Fig. 2E, blue**) compared to the standard condition ($p \ll 0.001$), which we interpret as being driven by the sMR costs in the absence of benefits of high gMR. In other words, mutation rate is selected against because of its somatic costs and the absence of benefits of higher gMR in static conditions. In natural populations that are under stabilizing selection, gMR will have costs due to greater phenotypic variance from a well-adapted state that are independent of sMR, but we do not model stabilizing selection in this study.

To investigate the role of the putative gMR benefit versus sMR cost balance in evolution, we further decoupled gMR and sMR by allowing gMR to evolve but making sMR cost fixed and independent of gMR (see **Methods: Model variations**). Decoupling sMR cost from gMR significantly accelerated the evolution of body size (**Fig. 2G, blue**) relative to the standard condition (green; $p = 0.0052$), revealing that sMR costs can limit the evolution of larger body size. During the early fast evolution of body mass, gMR (**Fig. 2H, blue**) and SMP (**Fig. 2I, blue**) demonstrate a corresponding positive response. Later, further body mass evolution becomes impeded (likely because of the severe depletion in population numbers), coinciding with selection against gMR. SMP plateaus during this second phase at a significantly lower level compared to the standard condition ($p \ll 0.001$), indicating that the somatic costs of mutation rate stimulate the evolution of more robust SMP.

As we have seen under blocked selection for body size (**Fig. 2B,C, blue**), SMP demonstrates an early phase of positive selection (**Fig. 2C, blue**) that is apparently reflected in a corresponding positive selection for gMR (**Fig. 2B, blue**). This observation suggests that both SMP and gMR may also respond to selection acting on some other traits, e.g. reproductive parameters

(**Supplements: Section 2**). This raises the question whether SMP and gMR evolution would be sensitive to strong selection for a trait that does not affect somatic risks (greater body size increases the target size for somatic mutations). We simulated a condition that was similar to the standard condition, except positive selection was applied to a trait that did not affect sMR related somatic costs (see **Methods: Model variations**); e.g. if SMP improvement is solely a response to the increased sMR cost imposed by larger body, selection for an sMR cost unrelated trait should not drive improvements in SMP. As shown in **Fig. 2J (blue)**, unimpeded by increased sMR costs and declining population size, the evolution of an sMR cost unrelated trait is significantly faster compared to the evolution of increased body size ($p \ll 0.001$). Interestingly, gMR (**Fig. 2K, blue**) also demonstrated an early phase of positive selection during early rapid evolution of the selected trait and remains above the initial gMR throughout the entire simulation. As anticipated, SMP is positively selected, however in the absence of an increasing sMR cost (associated with larger bodies), SMP's improvement is significantly smaller (**Fig. 2L, blue,** $p \ll 0.001$). Notably, even with much less enhanced SMP, gMR is still under positive selection in response to positive selection of the sMR cost unrelated trait (**Fig. 2L, blue**), consistent with the sMR/gMR cost/benefit ratio being an important factor regulating selection acting on gMR. Regardless, the results demonstrate that both gMR and SMP are responsive to selection for somatic risk unrelated traits, which indicates that high mutation rate is beneficial in positively selective conditions.

As we have seen in **Fig. 2D-F**, in the absence of strong positive selection for body size and SMP efficiency, selection acts to lower gMR. **Fig. 3** shows, however, that this selection is significantly modified by the efficiency of SMP. Stronger SMPs (lower *Som* value) relax selection for lower gMR when directional selection is weak (non-overlapping CIs between the standard (red) and either of the improved SMPs). As will be explained further below, this observation may have significant implication on long-term species survival.

Under strong positive selection, whether for body mass (**Fig. 1A-C, blue**) or a sMR cost unrelated trait (**Fig. 2H,I, blue**, and **Fig. 2K,L, blue**), gMR demonstrates consistent signs of positive selection. However, because gMR and sMR are linked, higher gMR is a trait that should

negatively impact individual fitness and therefore be under negative selection. To investigate this question, we mixed two simulated genotypes, one "wild-type" (50%) and one "mutator" (50%) in a population of stable size and under positive selection for a sMR cost unrelated trait. We then observed the genotypes' frequencies in the population using varying strength of mutators. **Fig. 4A** demonstrates that while the mutator's fitness initially is lower compared to wild-type, eventually the mutator outcompetes its wild-type counterpart. Interestingly, with increased mutation rate, the magnitude of the mutator's initial decline increases, but so does the speed at which it subsequently overtakes the population. This result provides a clue for how higher mutation rate, being a trait with negative impact on fitness, can be selected for. Because net organismal fitness is a composite trait impacted by the fitness value of many individual traits, the initial fitness of the "mutator" is lower because, all other traits equal, higher MR incurs increased sMR cost. However, in response to selection, mutator is capable of more rapidly developing other (adaptive) traits (**Fig. 4B**) and thus its overall fitness soon becomes higher compared to wild-type. Its noteworthy that genetic recombination in sexually reproducing populations should theoretically act to segregate adaptive alleles (under positive selection) from mutator alleles that are not directly selected for and even should be negatively selected. **Fig. 5** shows a model that we propose to explain how small population size should effectively impede such allelic segregation under positive selection. Importantly, **Fig. 5** also demonstrates that higher gMR is only beneficial under positive selection, while stabilizing selection will act to lower it even in the absence of the incumbent somatic risks.

Discussion

Our study demonstrates that positive selection for body size triggers a concurrent selection for improved somatic maintenance to mitigate the increased somatic risks of larger bodies. Improved somatic maintenance, in turn, promotes selection for higher germline mutation rates by reducing the cost of somatic mutations and thus altering the sMR/gMR cost/benefit ratio. Conditions of strong positive selection for other than SMP traits, as our model shows, can also

alter this balance by elevating the benefits of higher gMR. Under stable conditions, alternatively, the sMR/gMR cost/benefit balance is altered by the existing cost of somatic mutations and by the increased cost and absent/reduced benefits of gMR itself (as shown in **Fig. 5A**), which ultimately favors lower mutations rates. Under stasis, gMR exerts a cost independent of somatic risks by increasing deviation of progeny phenotypes from population mean/median and thus reducing their fitness. Our study thus demonstrates that the evolution of mutation rate is not exclusively limited by negative selection and population size, but is highly tunable and governed by selection acting on other traits. Importantly, our modeling indicates that under certain conditions elevated mutation rate, unlike perhaps any other trait, can be positively selected despite its negative effects on individual fitness (as explained in **Fig. 4**). Mutation rate, therefore, does not entirely fit in the paradigm formulated by George C. Williams (Williams, 1966) that evolution does not have eyes for the future (which appears universal for other traits). Being maladaptive in stable conditions, higher mutation rate becomes a trait that improves the net multi-trait fitness in conditions of positive selection for other traits by generating greater diversity of other traits, thus increasing a population's sensitivity to selection and accelerating adaptation. These observations can provide an explanation why mutation rate, although showing some major patterns, neither strictly follows phylogeny nor population size in mammals as shown by Lynch (Lynch, 2010).

Mutation rate in eukaryotes is a highly polygenic trait encoded by multiple genes involved in DNA replication, repair and cell division machineries (Pothof et al., 2003, Galetzka et al., 2007). Animals mostly reproduce sexually, which should generate an extensive population allelic diversity for these genes. This diversity should provide for a relatively continuous distribution of mutation rate in populations, rather than being a uniform trait marked with sporadic monogenic mutants, as may occur in asexual populations (Cox & Gibson, 1974, Gibson et al., 1970, Sniegowski et al., 1997). Such intra-population variation (Harris, 2015, Conrad et al., 2011), as well as the ability of mutation rate to rapidly evolve (Harris & Pritchard, 2017), has been shown for humans . However, sexual reproduction would be supposed to effectively segregate alleles contributing to mutation rate from alleles for other (e.g. adaptive) traits. It has been argued based

on other evidence that the efficiency of such segregation in sexual populations is limited (Draghi & Wagner, 2008). Here, we argue that given the polygenic nature of mutation rate, such segregation should be much less efficient in small populations that are under positive selection, and should be substantially impeded by selection for extreme phenotypes (as shown in **Fig. 5**). The polygenic nature of mutation rate should also impede segregation of mutator phenotypes from adaptive phenotypes, as most genes contributing to the overall mutation rate will individually have rather modest effects on fitness and in many cases their effect on fitness may depend on the allelic composition of other loci. In monogenic traits, on the other hand, a single locus will have a defined effect on the net phenotype and thus will directly affect selection acting on it.

It also appears from our results that animal evolution, with the macroscopic trend toward larger bodies, should have driven a concurrent evolution of extended longevity, the latter being determined by the efficiency of species-specific somatic maintenance programs. Even though extended longevity tentatively appears to be a benefit on its own, e.g. due to extended reproduction period, our model demonstrates that somatic maintenance (and thus longevity) is under a much weaker positive selection in the absence of other positively selected traits. This observation can explain why extended longevity demonstrates significant deviations across animal taxa from the general rule larger body → longer lifespan. Our results indicate that the evolution of longevity (as a function of somatic maintenance efficiency) should be greatly impacted by the rate of evolution of other traits, and not necessarily body size.

Interestingly, our study predicts an important evolutionary role for the mechanisms of somatic maintenance in addition to their evolution as a means of improving individual survival of large animals (Caulin & Maley, 2011, Rozhok & DeGregori, 2016). Our results demonstrate that selection for enhanced somatic maintenance goes well beyond the evolution of body size and is promoted by strong directional selection acting on any trait. This result indicates that SMPs may have had an important role in the evolution of large animals. Selection for higher gMR ensuing improved SMP may be an important mechanism “rescuing” the reduced evolvability imposed by reduced population size, extended generation times and lower reproduction rates. Therefore,

SMPs and longevity may have an important contribution to species long-term survival. For example, a prolonged evolutionary stasis (Benton & Pearson, 2001, Eldredge & Gould, 1972, Gould & Eldredge, 1993, Venditti et al., 2011) should trigger selection for lower mutation rates. By relaxing negative selection on mutation rate and thus maintaining evolvability (as shown in **Fig. 3**), enhanced SMPs can ensure better survival of animal groups facing rapid evolutionary transitions or drastically changed environments after such relatively static periods. All other traits equal, species with extended longevity may survive such transitions with higher probabilities.

Lynch and colleagues have provided extensive arguments supporting the idea that the higher MRs in animals compared to unicellular organisms are likely to be caused by reduced population sizes that limit the ability of negative selection to act on mutation rate (Lynch, 2010, Lynch, 2011, Lynch et al., 2016). In conjunction with population size, in large animals the strength of selection will be further attenuated by lower reproduction rates and extended generation times. Based on our results, Lynch's theory can be extended by recognizing that somatic maintenance programs (and longevity) should have substantial influence on the general relationship between population size and mutation rates, and on the strength and directionality of selection acting on mutation rates. For example, in our simulation, populations of the same initial size but with different SMP efficiencies demonstrate profound differences in the effects of population size driven weakening of selection (**Fig. 1F,G**, as well as discrepant selection for mutation rates (**Fig. 1D**).

Selection for higher mutation rates has been shown experimentally in bacteria (Gibson et al., 1970, Cox & Gibson, 1974, Sniegowski et al., 1997, Loh et al., 2010), whereby engineered or spontaneous mutants with higher mutation rate have been shown to have advantages over wild-type in positively selective conditions. The "mutator hitchhiker hypothesis" explains such selection by the higher probability that adaptive mutations will appear in a mutator cell (Sniegowski et al., 1997). Once such a mutation occurs, the mutator genotype spreads to fixation by being genetically linked to the adaptive phenotype. Modeling studies demonstrate that evolution of evolvability, including varying selection on mutation rates, should be possible in

sexually reproducing organisms (Jones et al., 2014, Draghi & Wagner, 2008, Jones et al., 2007). Yet robust experimental corroboration of such a possibility appears to be lacking.

In conclusion, our results raise the question of whether the evolution of large body size in animals would be possible without such a complex pattern of selection acting on mutation rate, and whether such a complex relationship is necessary to explain the evolution of large animals. The evolution of large bodies has entailed the cost of losing the ability to evolve via all major parameters that define this ability, such as population size, reproduction rate and generation time, <u>except</u> mutation rate (which increased). Therefore, one scenario could have been that this cost has been so prohibitive for many species that positive selection for mutation rate was necessary to allow evolution of large animals. Alternatively, mutation rate could have been high enough to maintain evolvability at the negative selection/drift barrier point where negative selection was no longer able to reduce it further (Lynch, 2010). Understanding which of these scenarios prevails in the evolution of large animals requires more research.

References

Baer, C. F., Miyamoto, M. M. & Denver, D. R. 2007. Mutation rate variation in multicellular eukaryotes: causes and consequences. *Nat Rev Genet* **8**: 619-31.

Baker, J., Meade, A., Pagel, M. & Venditti, C. 2015. Adaptive evolution toward larger size in mammals. *Proc Natl Acad Sci U S A* **112**: 5093-8.

Benton, M. J. & Pearson, P. N. 2001. Speciation in the fossil record. *Trends Ecol Evol* **16**: 405-411.

Caulin, A. F. & Maley, C. C. 2011. Peto's Paradox: evolution's prescription for cancer prevention. *Trends Ecol Evol* **26**: 175-82.

Conrad, D. F., Keebler, J. E., DePristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C. L., Torroja, C., Garimella, K. V., Zilversmit, M., Cartwright, R., Rouleau, G. A., Daly, M., Stone, E. A., Hurles, M. E. & Awadalla, P. 2011. Variation in genome-wide mutation rates within and between human families. *Nat Genet* **43**: 712-4.

Cox, E. C. & Gibson, T. C. 1974. Selection for high mutation rates in chemostats. *Genetics* **77**: 169-84.

Dawson, K. J. 1999. The dynamics of infinitesimally rare alleles, applied to the evolution of mutation rates and the expression of deleterious mutations. *Theor Popul Biol* **55**: 1-22.

Draghi, J. & Wagner, G. P. 2008. Evolution of evolvability in a developmental model. *Evolution* **62**: 301-15.

Eldredge, N. & Gould, S. J. (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: *Models in Paleobiology*, (Schopf, T. J. M., ed.). pp. 82-115. Freeman Cooper, San Francisco.

Finette, B. A., Sullivan, L. M., O'Neill, J. P., Nicklas, J. A., Vacek, P. M. & Albertini, R. J. 1994. Determination of hprt mutant frequencies in T-lymphocytes from a healthy pediatric population: statistical comparison between newborn, children and adult mutant frequencies, cloning efficiency and age. *Mutat Res* **308**: 223-31.

Galetzka, D., Weis, E., Kohlschmidt, N., Bitz, O., Stein, R. & Haaf, T. 2007. Expression of somatic DNA repair genes in human testes. *J Cell Biochem* **100**: 1232-9.

Gibson, T. C., Scheppe, M. L. & Cox, E. C. 1970. Fitness of an Escherichia coli mutator gene. *Science* **169**: 686-8.

Giese, H., Snyder, W. K., van Oostrom, C., van Steeg, H., Dolle, M. E. & Vijg, J. 2002. Age-related mutation accumulation at a lacZ reporter locus in normal and tumor tissues of Trp53-deficient mice. *Mutat Res* **514**: 153-63.

Glick, D., Barth, S. & Macleod, K. F. 2010. Autophagy: cellular and molecular mechanisms. *J Pathol* **221**: 3-12.

Gould, S. J. & Eldredge, N. 1993. Punctuated equilibrium comes of age. *Nature* **366**: 223-7.

Hanahan, D. & Weinberg, R. A. 2000. The hallmarks of cancer. *Cell* **100**: 57-70.

Harris, K. 2015. Evidence for recent, population-specific evolution of the human mutation rate. *Proc Natl Acad Sci U S A* **112**: 3439-44.

Harris, K. & Pritchard, J. K. 2017. Rapid evolution of the human mutation spectrum. *Elife* **6**.

Heim, N. A., Knope, M. L., Schaal, E. K., Wang, S. C. & Payne, J. L. 2015. Animal evolution. Cope's rule in the evolution of marine animals. *Science* **347**: 867-70.

Hochberg, M. E. & Noble, R. J. 2017. A framework for how environment contributes to cancer risk. *Ecol Lett* **20**: 117-134.

Horvath, S. 2013. DNA methylation age of human tissues and cell types. *Genome Biol* **14**: R115.

Jones, A. G., Arnold, S. J. & Burger, R. 2007. The mutation matrix and the evolution of evolvability. *Evolution* **61**: 727-45.

Jones, A. G., Burger, R. & Arnold, S. J. 2014. Epistasis and natural selection shape the mutational architecture of complex traits. *Nat Commun* **5**: 3709.

Kennedy, S. R., Loeb, L. A. & Herr, A. J. 2012. Somatic mutations in aging, cancer and neurodegeneration. *Mech Ageing Dev* **133**: 118-26.

Kimura, M. 1967. On the evolutionary adjustment of spontaneous mutation rates. *Genetical Research* **9**: 23-34.

Loh, E., Salk, J. J. & Loeb, L. A. 2010. Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc Natl Acad Sci U S A* **107**: 1154-9.

Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. 2013. The hallmarks of aging. *Cell* **153**: 1194-217.

Lotka, A. J. 1925. *Elements of Physical Biology*. Williams and Wilkins.

Lynch, M. 2010. Evolution of the mutation rate. *Trends Genet* **26**: 345-52.

Lynch, M. 2011. The lower bound to the evolution of mutation rates. *Genome Biol Evol* **3**: 1107-18.

Lynch, M., Ackerman, M. S., Gout, J. F., Long, H., Sung, W., Thomas, W. K. & Foster, P. L. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet* **17**: 704-714.

Madsen, T., Arnal, A., Vittecoq, M., Bernex, F., Abadie, J., Labrut, S., Garcia, D., Faugère, D., Lemberger, K. & Beckmann, C. 2017. Cancer Prevalence and Etiology in Wild and Captive Animals. *Ecology and Evolution of Cancer*: 11.

Marcon, E. & Moens, P. B. 2005. The evolution of meiosis: recruitment and modification of somatic DNA-repair proteins. *Bioessays* **27**: 795-808.

Pfau, S. J., Silberman, R. E., Knouse, K. A. & Amon, A. 2016. Aneuploidy impairs hematopoietic stem cell fitness and is selected against in regenerating tissues in vivo. *Genes Dev* **30**: 1395-408.

Pothof, J., van Haaften, G., Thijssen, K., Kamath, R. S., Fraser, A. G., Ahringer, J., Plasterk, R. H. & Tijsterman, M. 2003. Identification of genes that protect the C. elegans genome against mutations by genome-wide RNAi. *Genes Dev* **17**: 443-8.

Rozhok, A. I. & DeGregori, J. 2016. The evolution of lifespan and age-dependent cancer risk. *Trends Cancer* **2**: 552-560.

Sherr, C. J. 2004. Principles of tumor suppression. *Cell* **116**: 235-46.

Sniegowski, P. D., Gerrish, P. J. & Lenski, R. E. 1997. Evolution of high mutation rates in experimental populations of E. coli. *Nature* **387**: 703-5.

Swann, J. B. & Smyth, M. J. 2007. Immune surveillance of tumors. *J Clin Invest* **117**: 1137-46.

Venditti, C., Meade, A. & Pagel, M. 2011. Multiple routes to mammalian diversity. *Nature* **479**: 393-6.

Volterra, V. 1926. Variazioni e fluttuazioni del numero d'individui in specie animali conviventi. *Mem Acad Lincei Roma* **2**: 31-113.

Williams, G. C. 1966. *Adaptation and Natural Selection: A Critique of Some Current Evolutionary Thought*. Princeton University Press.

Figure legends

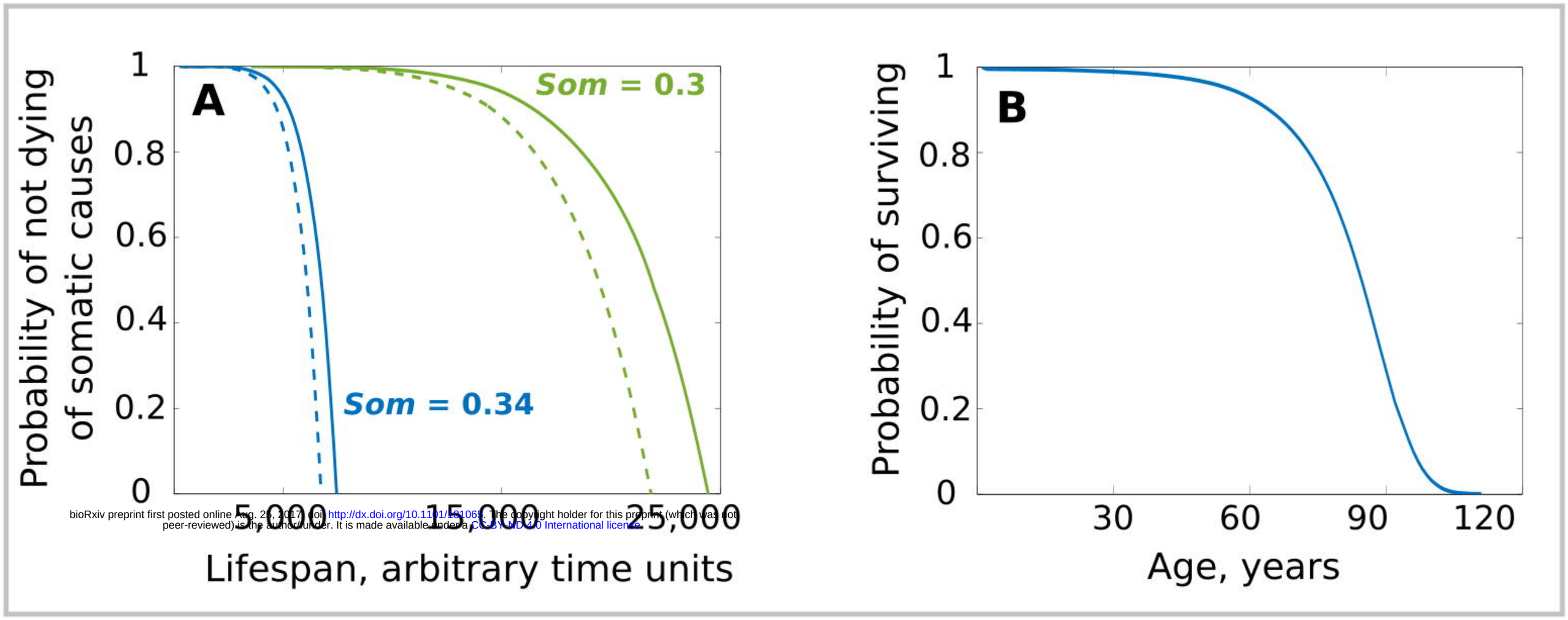
Fig. 1. The effect of SMP evolution on the evolution of body mass and mutation rate. (*A*) physiological/aging related mortality curves generated based on the cumulative distribution function of D_A (Eq. 1). Colors represent the effect of the *Som* (SMP) parameter (Eq. 1). Dotted lines were generated by elevating mutation rate 2-fold. (*B*) modern human mortality in the U.S.A (https://www.ssa.gov). (*C*) evolution of life history traits under positive selection for body size. (*F,G*) population size dynamics when SMP can evolve (corresponds to green in *C-E*) or SMP evolution is blocked (blue in *C-E*); colors indicate individual populations. (*H*) relative frequency of Species B (SMP evolution blocked, blue in *C-E*) in a mixed population with Species A (SMP can evolve, green in *C-E*). For (*C*), (*D*), (*E*) and (*H*) (and similar graphs in other figures), 25 simulations are combined, with the dark line reflecting the mean and shaded area denoting the 95% confidence intervals.

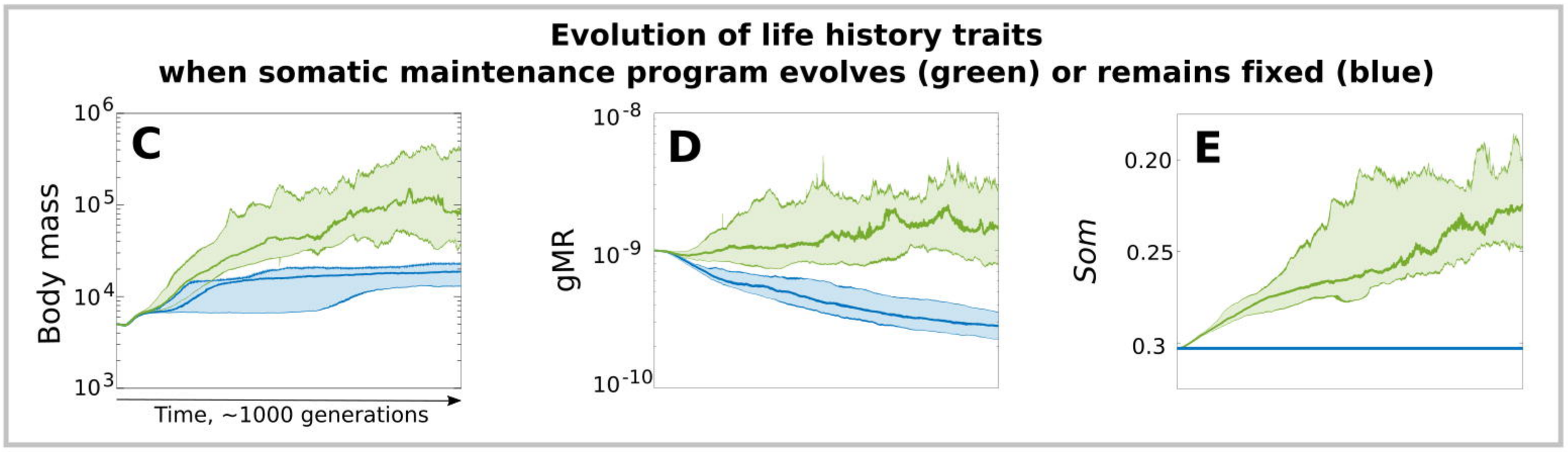
Fig. 2. Evolution of body mass, gMR and SMP under various regimens of selection. Separate experiments are stacked as indicated in their subtitles. The layout: left – body size, middle – gMR, right – SMP (the *Som* parameter in Eq. 1) is maintained as in Fig. 1C-E. Green – the standard condition (as green in Fig. 1C-E); blue – alternative conditions with fixed values of a trait (blue horizontal line in *A,D,F*), when gMR and sMR are dislinked so that the somatic cost is fixed while gMR can evolve (blue in G-I) and under selection for a somatic risk unrelated trait (blue in *J-L*).

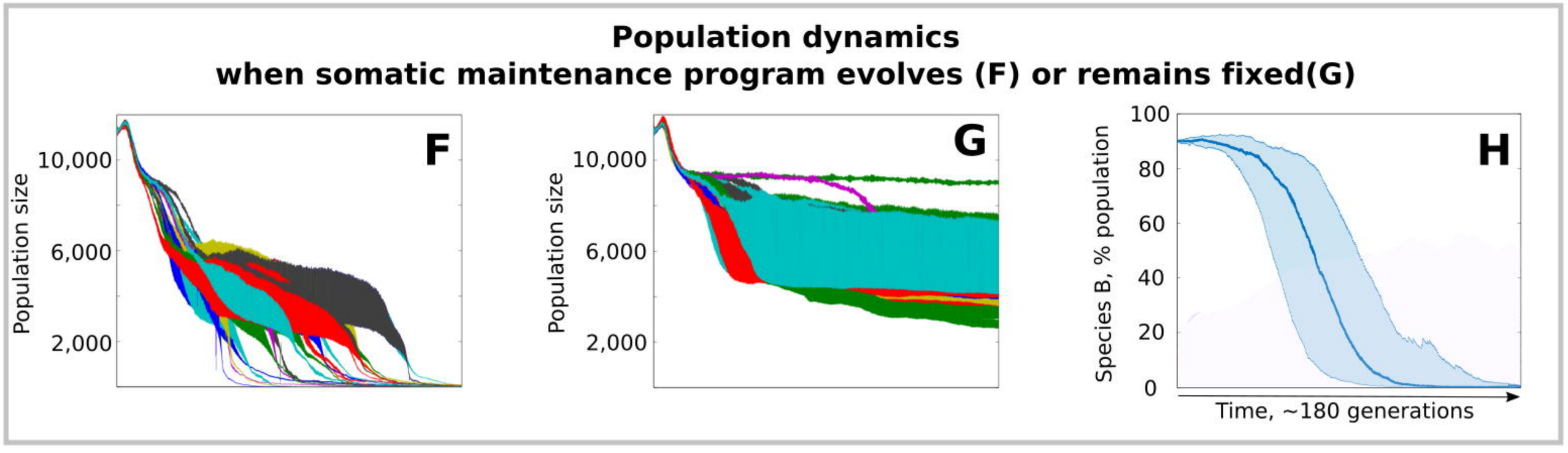
Fig. 3. The evolution of gMR in in the absence of positive selection for body mass and SMP. The SMP's *Som* parameter was fixed at 0.34 (red), 0.24 (green; enhanced 10X) and 0.2 (blue; enhanced 40X); a linear decrease in the *Som* value results in a substantially improved SMP, so that the green SMP is ~10X more efficient compared to red, and the blue is a ~4X more efficient SMP than the green. The standard (red) SMP leads to a significantly stronger selection for lower gMR (non-overlapping 95% CIs); however, the absence of difference between the 10X (green) and 40X (blue) improved SMPs indicates that overly improved SMPs might not provide any further difference for how selection acts on gMR.

Fig. 4. Positive selection for mutators. (*A*) frequency of a mutator phenotype in a mixed competitive population with "wild-type" species. Red (1.4X), orange (2X) and green (10X) are mutators of different fold increase in MR relative to the competitor as indicated by the respective numbers. (*B*) positive selection for a somatic cost neutral trait demonstrates faster evolution (and so adaptation) of mutators. Colors and MR fold increase as in (*A*).

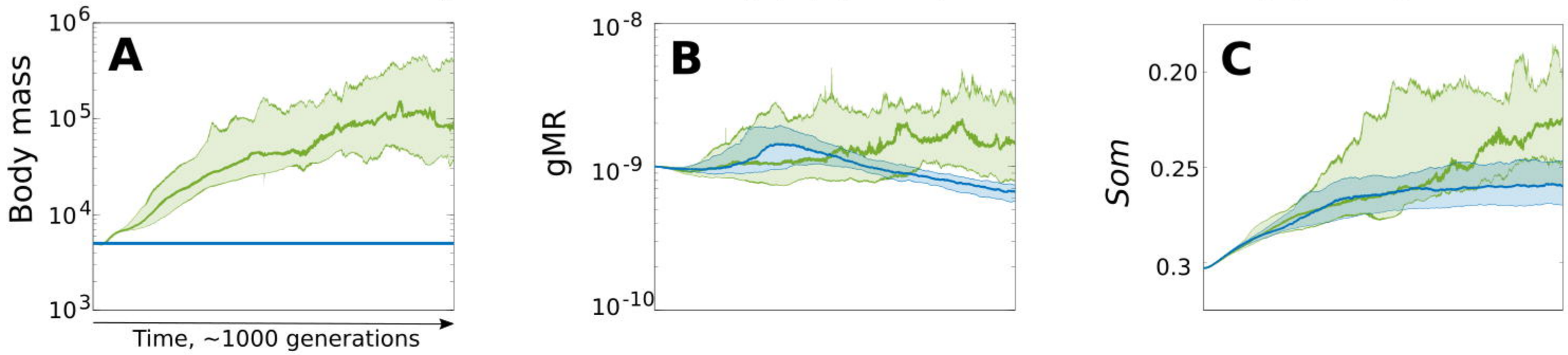
Fig. 5. A model of how selection acts on mutation rate in sexual populations. (*A*) under stabilizing selection, the most adaptive phenotypes are close to the population mean/median; such phenotypes are more likely to be produced by parents with low germline mutation rate in a population in which mutation rate is a multi-genic distributed trait. (*B*) under positive selection, the most adaptive phenotypes demonstrate unidirectional deviation of the selected trait(s) from the population mean. Such phenotypes are more likely to be produced by parents having higher germline mutation rate and thus harboring multiple alleles conducive to higher mutation rate; (*C*) small population size reduces the strength of selection by increasing the strength of drift; this condition requires a phenotype to deviate sufficiently far from the population mean/median towards the selected tail to be responsive to selection. Such extremely deviant phenotypes in small populations are likely to come from parents with the highest germline mutation rate and thus harboring fewer alleles for low mutation rate. This condition should impede segregation of mutator alleles from adaptive alleles by recombination imposed by sexual reproduction.



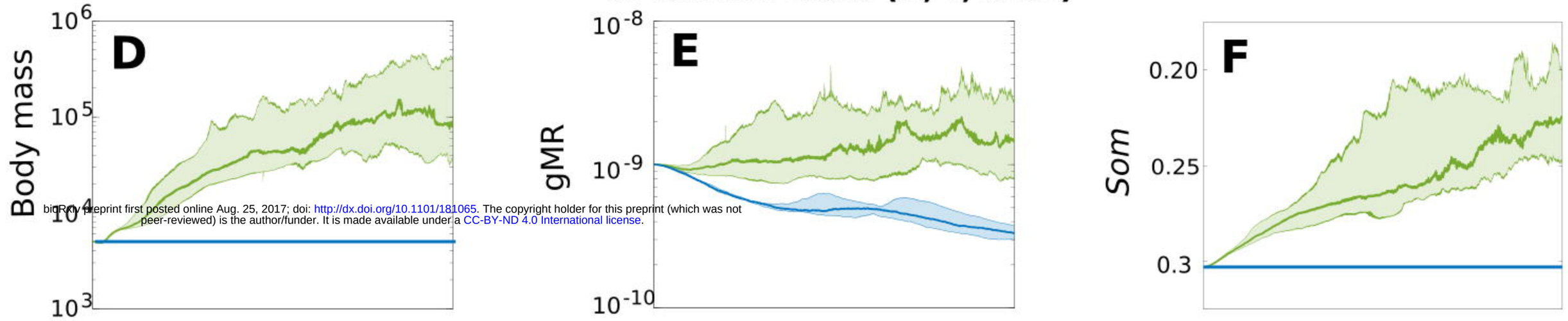




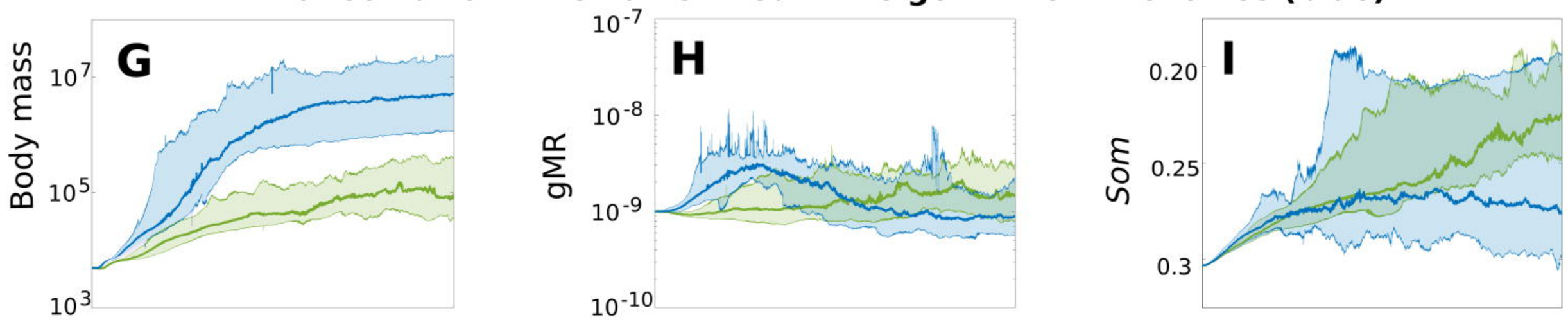
Evolution of mutation rate (B) and somatic maintenance (C) when body mass is evolving (A, green) or remains fixed (A, blue)



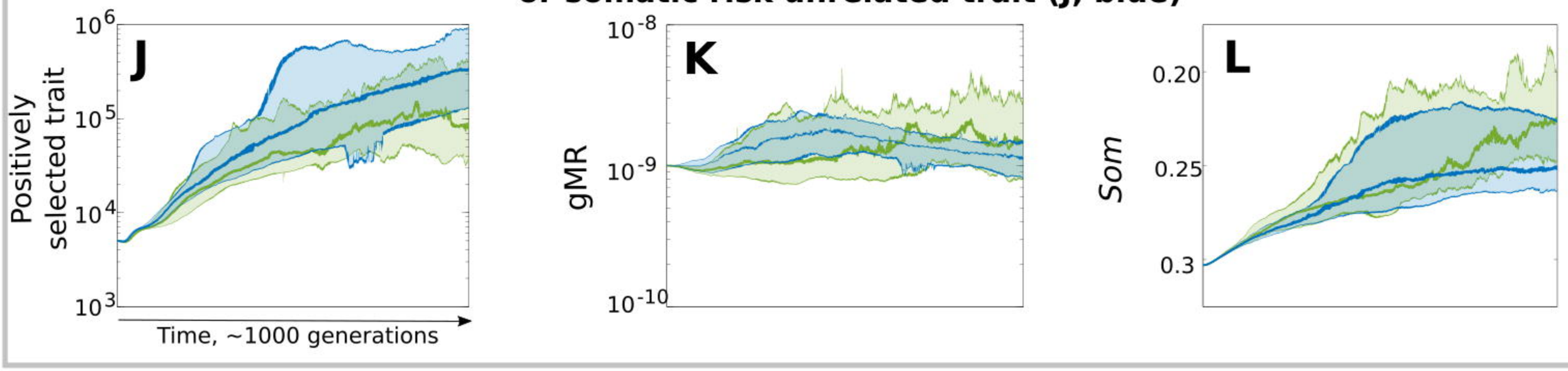
Evolution of mutation rate (E) when body mass and somatic maintenance are evolving (A, F, green) or remain fixed (A, F, blue)

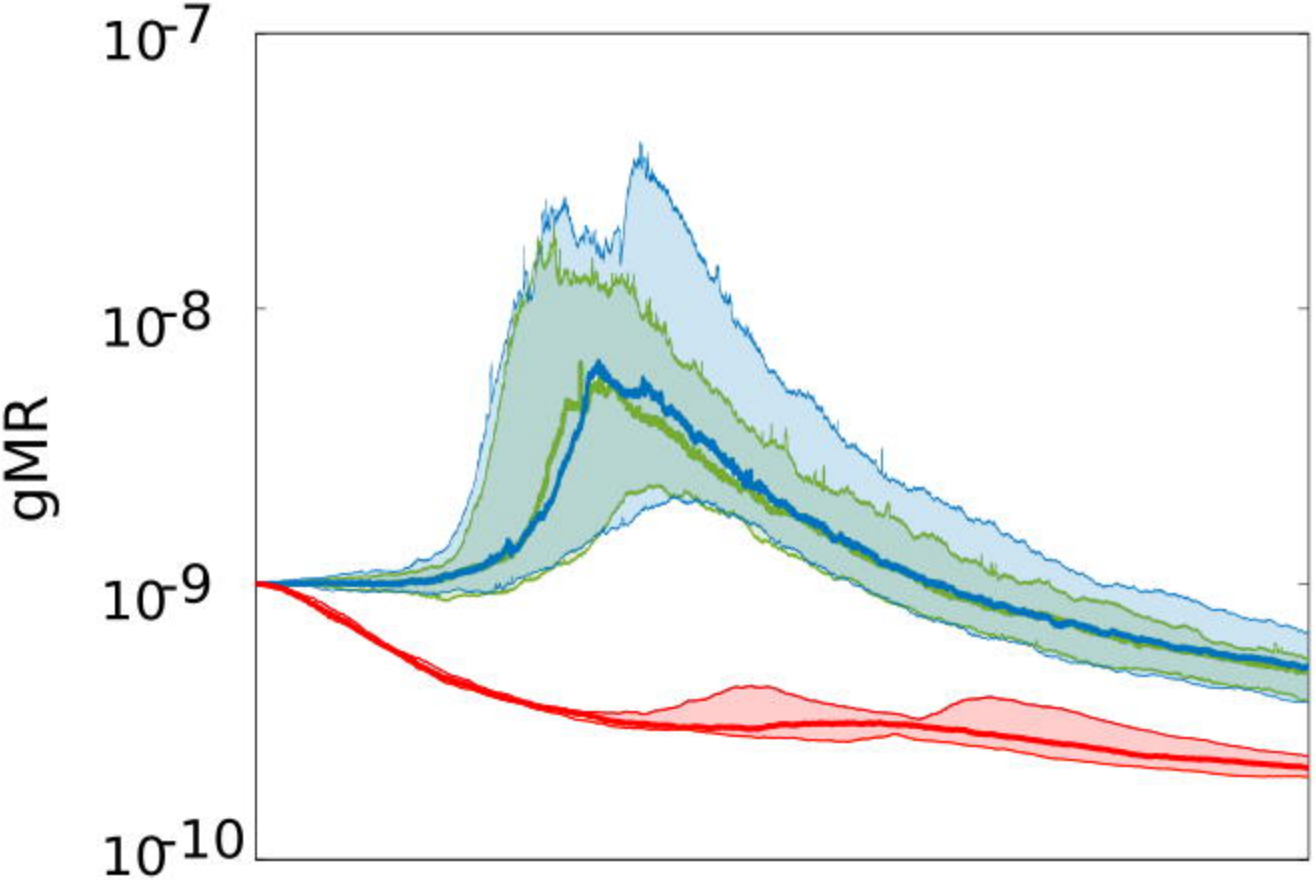


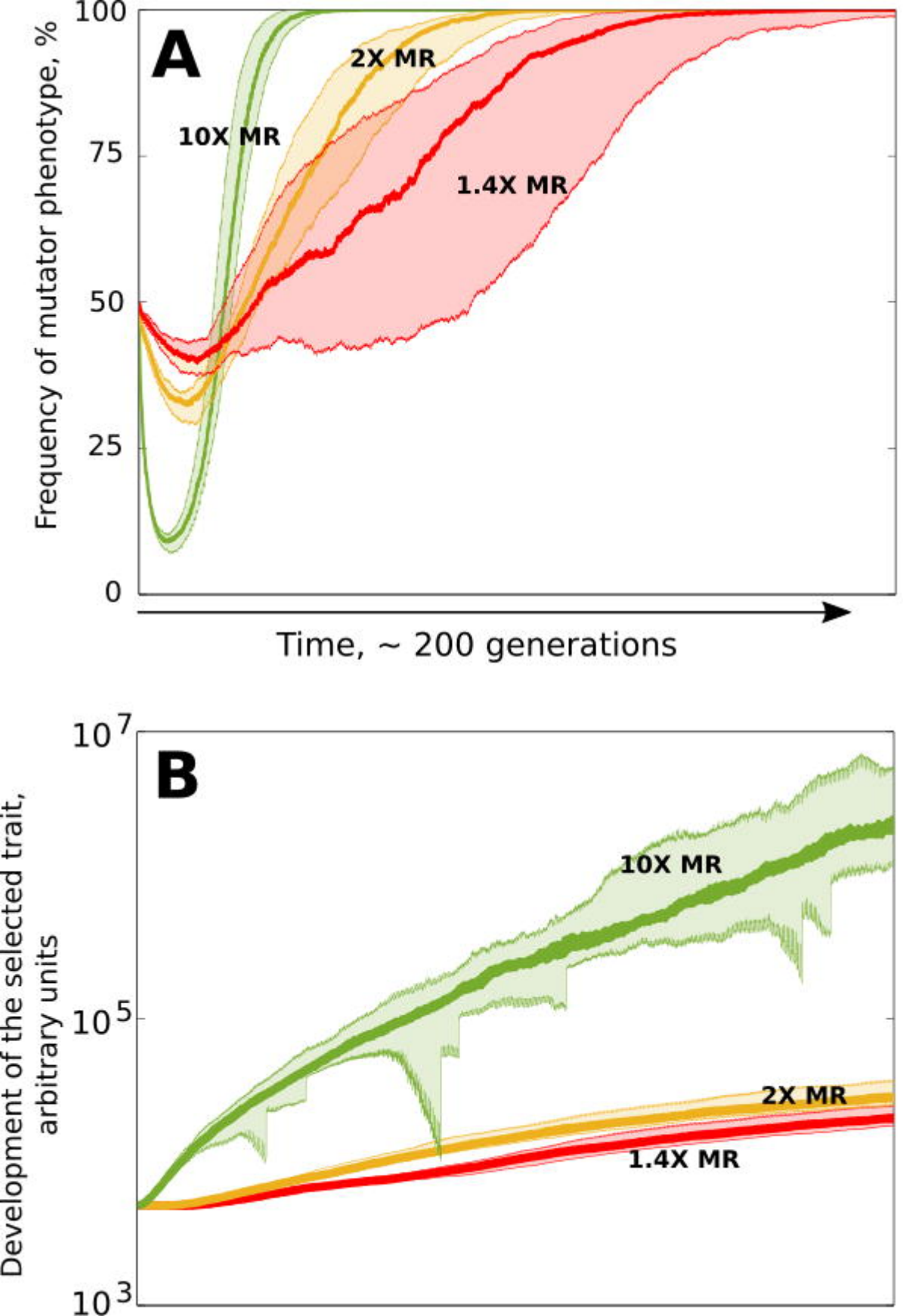
Evolution of life history traits when somatic and germline mutation rates, MR, are linked (green) or somatic MR remains fixed while germline MR evolves (blue)

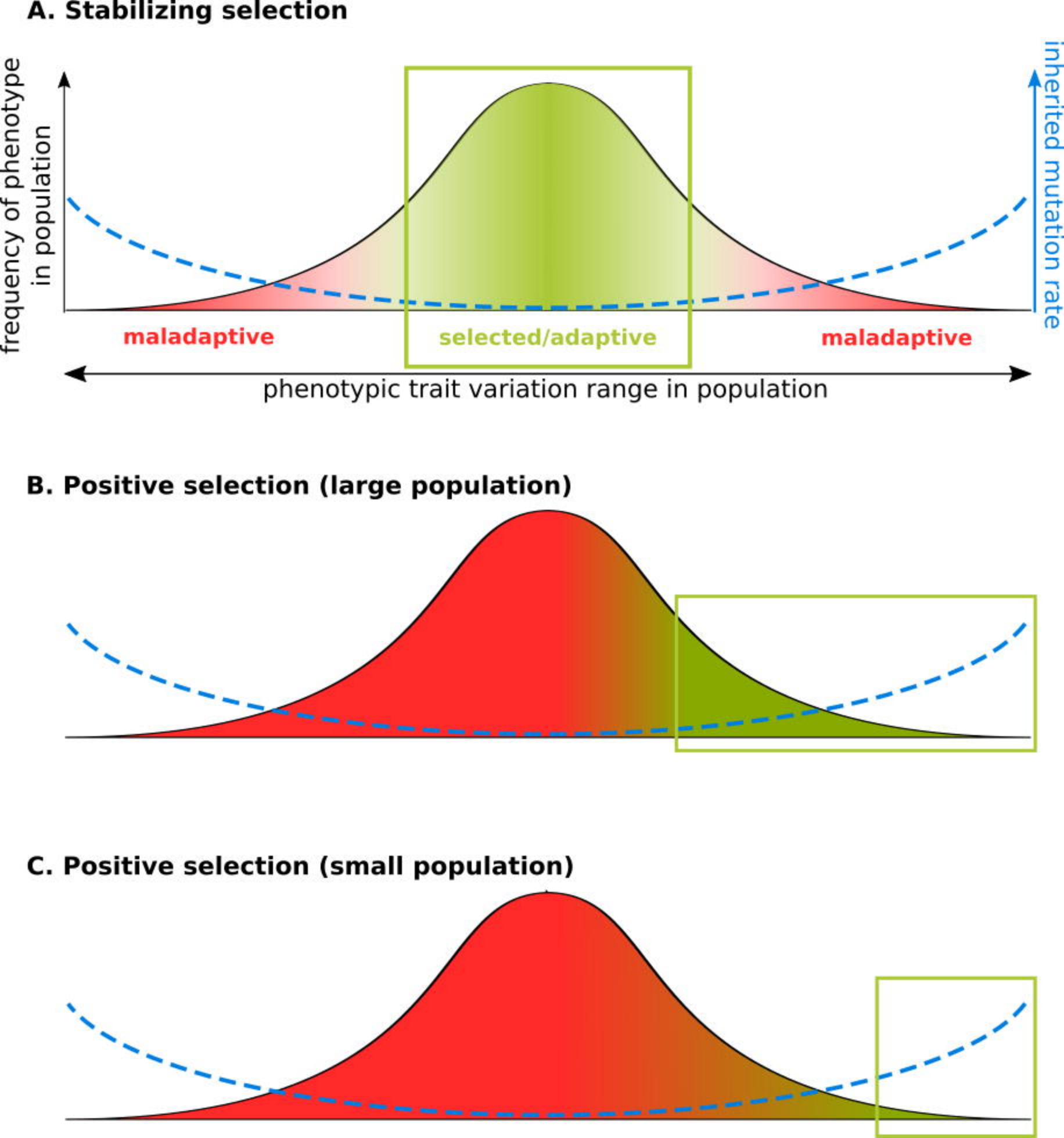


Evolution of mutation rate (K) and somatic maintenance (L) under positive selection for body size (J, green) or somatic risk unrelated trait (J, blue)




10^{-7}
10^{-8}
10^{-9}
10^{-10}
gMR


A
100
75
50
25
0
Frequency of mutator phenotype, %
10X MR
2X MR
1.4X MR
Time, ~ 200 generations
B
10^7
10^5
10^3
Development of the selected trait, arbitrary units
10X MR
2X MR
1.4X MR


A. Stabilizing selection
frequency of phenotype in population
inherited mutation rate
maladaptive
selected/adaptive
maladaptive
phenotypic trait variation range in population
B. Positive selection (large population)
C. Positive selection (small population)