Preprints.org

Hypothesis

Not peer-reviewed version

The Underappreciated Role of Reproductively Lethal Mutations in the Evolution of Living Beings

Feng Qu * , Khwannarin Khemsom , Camila Perdoncini Carvalho , Junping han

Posted Date: 29 September 2023

doi: 10.20944/preprints202309.2051.v1

Keywords: Lethal mutation; reproductively lethal mutation; reproductive bottlenecking; evolution; natural selection; phenotype; asexual reproduction; sexual reproduction; virus; bacterium; plant; insect, animal; eusociality

Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

The Underappreciated Role of Reproductively Lethal Mutations in the Evolution of Living Beings

Feng Qu *, Khwannarin Khemsom, Camila Perdoncini Carvalho and Junping Han

Department of Plant Pathology, The Ohio State University, Wooster, OH 44691; qu.28@osu.edu

Abstract: Reproductively lethal mutations (RLMs) are mutations that, upon expression of the encoded lethal phenotypes, cause individuals carrying them to die or to be sterile. An underappreciated fact is that loss-of-function RLMs in protein-coding genes can be phenotypically shielded by favorable environments rendering these genes conditionally non-essential, or by their sister alleles in diploid organisms. Absent of rigorous mitigation, such phenotype-shielding causes the number of genes incurring RLMs to increase over time, and simultaneously allows each RLM to reach high allele frequencies in a conspecific population. Over-accumulation of RLMs then sets the population up for eventual concurrent expression of large numbers of RLMs, and massive deaths in rapid succession, possibly even population-level extinction. This hypothetical scenario in turn predicts that organismal lineages that evolved means to minimize the allele frequencies of phenotypically shielded RLMs are favored by natural selection. We argue that bottlenecking the genome copies destined for reproduction is a universal strategy adopted by all living beings to compel phenotype-based RLM purging. We further postulate that primitive RNA replicons must first evolve bottlenecked reproduction before evolving the capacity to encode diffusible products. In more complex, multicellular organisms, RLM management through bottlenecked reproduction gains additional reinforcement through sexual reproduction. In short, the evidence chronicled in this essay strongly suggest that the Bottleneck, Isolate, Amplify, Select (BIAS) principle, originally proposed to explain intracellular evolutionary dynamics of viruses, may be universally applicable to all living beings.

Keywords: lethal mutation; reproductively lethal mutation; reproductive bottlenecking; evolution; natural selection; phenotype; asexual reproduction; sexual reproduction; virus; bacterium; plant; insect; animal; eusociality

Introduction

A lethal mutation is a mutation that, when given the chance to express the encoded lethality, causes the individual organism carrying this mutation to die. Such individuals, especially in humans, attract most attention when they die shortly after birth. Less recognized are perhaps many more deaths that occur earlier during pregnancy, due to the expression of lethal phenotypes at various stages of prenatal development A lethal mutation is a mutation that, when given the chance to express the encoded lethality, causes the individual organism carrying this mutation to die. Such individuals, especially in humans, attract most attention when they die shortly after birth. Less recognized are perhaps many more deaths that occur earlier during pregnancy, due to the expression of lethal phenotypes at various stages of prenatal development (Amorim et al., 2017; Meinke, 2020; O'Rourke et al., 2011). In extreme cases, arrest of embryo development occurs immediately following fertilization, upon expression of lethal mutations in genes essential for cell viability, such as those encoding certain DNA-dependent DNA polymerases, RNA polymerases, ribosomal proteins, or regulators of cell division. At the opposite end of spectrum are mutation-carrying individuals who survive for substantial lengths of time after birth, but fail to yield descendants due to mutation-encoded sterility. Such mutations are considered lethal in the reproductive sense. Viewed from the evolutionary lens, the various types of lethal mutations described above would all be excluded from future generations as soon as they gain the chance to express their lethal phenotypes. In the current essay, we invoke the term reproductively lethal mutations (RLMs) to encompass all mutations whose expressed phenotypes preclude the carriers from spreading the mutations in the population/species to which the carriers belong.

© 2023 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

RLMs that nullify essential genes are obligate RLMs. However, most RLMs are likely to be conditionally lethal, meaning the encoded lethality manifests only under certain conditions. For example, absent of insulin treatment, mutations that predispose Type 1 diabetes in humans are known to be conditionally lethal because their chance of causing life-threatening diabetes is greatly influenced by environmental triggers (Redondo et al., 2018). Similarly, without surgical intervention, genes that cause fetuses to develop larger heads could be lethal when carried by mothers with narrower pelvic canals, even if the fetuses were otherwise perfectly normal (Stålberg et al., 2006). Thus, a paradoxical consequence of modern medicine is to weaken purifying selection against certain conditional RLMs, thereby elevating their population-wide allele frequencies. This may in the long run adversely impact the rigor of human populations.

For molecular biologists, perhaps the most frequently encountered example of a conditional lethal phenotype is the death of commonly used lab strains of *E. coli* (e.g. DH5α) in media containing certain antibiotics. Such death is conditional because it occurs only when the media contain specific antibiotics. Conversely, in antibiotics-free medium these *E. coli* strains are fully capable of reproduction. Expanding from this example, some bacteria, such as the plant-pathogenic *Pseudomonas syringae*, can lose the ability to colonize plants upon extended propagation in culture media (Xin et al., 2018). This is because spontaneous mutations in *P. syringae* genes specifically required for plant colonization are conditional – they have little impact on bacterial reproduction in culture media, but are forced to unveil their conditional lethality upon encountering host plants. Further extrapolating from this example, it is not hard to imagine that within a relatively constant environment or ecological niche, most organisms are prone to conditionally lethal mutations that will only reveal their lethal consequences upon encountering drastic changes of the environment that necessitate the functionality of underlying genes.

We wish to emphasize that conditional RLMs should not be confused with deleterious, but nonlethal mutations. Conditional RLMs by definition cause the carriers to die or be sterile under conditions that compel the expression of the encoded lethal phenotypes, thereby excluding themselves from the conspecific population. By contrast, most nonlethal deleterious mutations merely cause the carriers to be phenotypically less competitive in reproduction. At the genomic DNA level, most conditional RLMs are loss-of-function mutations in protein-coding genes, caused by premature stop codons, transposon insertion, or other gene-disruption events. By contrast, most nonlethal deleterious mutations are probably nucleotide substitutions that partially hamper the function of the underlying genes.

RLMs inevitably emerge through genome replication. This is true for all organisms no matter how faithful their replication enzymes are (Ganai and Johansson, 2016; Lynch, 2010). A common misconception is that RLMs exert little impact in organisms equipped with a highly effective proof-reading system. However, such proof-reading systems, while greatly increasing the fidelity of genome replication, are incapable of eradicating all replication errors (Kunkel, 2009; Lynch, 2010; Lynch et al., 2016). More crucially, whether an error causes the carrier organism to die is unknowable until the encoded phenotypic perturbation(s) gain the chance to manifest. RLMs can also arise from other mutagenic events such as transposon transposition, erroneous meiotic crossover, or action of environmental mutagens. It is not the intention of current essay to review the many factors affecting the fidelity of genome replication processes in diverse organisms. It suffices to recognize that mutations occur in the genomes of all life forms, and some of the mutations will inevitably have lethal consequences when given the chance to express their lethality or sterility (Avenarius et al., 2009). This is vividly evidenced by recent reports that profiled mutations in individual sperms of human males (Arnheim and Calabrese, 2016; Bell et al., 2020; Goldmann et al., 2019). One such study concluded that "clonal mosaicism (meaning variations incurred during early steps of spermatogenesis that were inherited by lineages of sperm cells) likely contributes a transmissible, predicted pathogenic exonic variant for 1 in 15 men, representing a life-long threat of transmission for these individuals and a significant burden on human population health." (Yang et al., 2021). There is no doubt some of the pathogenic variants will have lethal consequences if given the chance to express their phenotypes, yet can spread unnoticed for generations if not expressed (more later).

In this essay, we argue that RLMs that are unable to express their phenotypes at the time of their emergence, such as conditional and recessive RLMs, pose grave danger to the long-term survival of the underlying organismal population or species. In the absence of rigorous surveillance that keep

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

these RLMs at low population-level allele frequencies, their accumulation over time sows the seed for massive deaths in rapid succession in conspecific populations. We further propose that bottlenecked reproduction, coupled with sexual reproduction in most multicellular species, is a universal mechanism that serves the role of minimizing the allele frequencies of conditional and recessive RLMs, primarily by forcing their phenotypic manifestation on a regular basis.

Absent of Constant Mitigation, Conditional and Recessive RLMs Can Accumulate to Levels that Endanger Species Survival

One might intuitively assume that RLMs, obligate or conditional, are of little evolutionary consequence, as failure of their carriers to parent live descendants automatically clears such mutations from the gene pool. However, it is worth repeating that RLMs must express their encoded lethality – meaning death or sterility of their carriers – to be excluded from descendant gene pools. Thus, RLMs that only manifest lethality under certain environmental conditions or genetic contexts can persist in the carrier genomes for many generations. Conditional RLMs are likely common for organisms that have short reproduction cycles, primarily because rapid reproduction correlates with more frequent genome replication and thus error introduction. On the other hand, such organisms typically have small body sizes and limited mobilities, making it difficult for them to evade adverse environments through migration. As a result, in these organisms the frequencies of conditional RLMs probably rise and fall in rapid successions in response to environmental swings that may seem to be modest to human standards.

Conversely, organisms with longer reproduction cycles, such as vertebrate animals and flowering plants, may incur fewer inheritable replication errors by limiting the number of cell divisions for dedicated germlines, hence also fewer RLMs. Additionally, individuals of these species also typically have longer life expectancies, giving them more time to experience broader ranges of environments, and evolve strategies to cope with fluctuations. They could also evade extreme conditions through migration. However, one must recognize that these larger organisms are also more capable of shielding conditional RLMs that do arise, for many more generations, permitting their spread in the populations/species, sowing seed for their concomitant expression in large numbers of individuals when faced with sudden environmental upheavals.

RLMs can also be shielded from purifying selection by diploid or polyploid genome arrangements. In diploid or polyploid organisms, RLMs that cause loss-of-function changes in protein-coding genes, regardless of their conditionality, are rarely purged immediately. This is because these recessive RLMs are shielded by their functional sister alleles from immediately expressing the encoded lethality, hence are transmitted to descendant genomes unnoticed. Put differently, recessive RLMs are not expunged from the carrier genomes until their sister alleles acquire the same or similar loss-of-function mutations (aka being homozygous). Such homozygosity can only be achieved through mating of two individuals carrying the same mutation (more later), or spontaneous loss-of-function mutations in the sister alleles.

Absent of active mitigation, persistence of recessive RLMs in diploids and polyploids poses serious threat to the long-term viability of these organisms. There is no denying that phenotypic masking of recessive RLMs can be beneficial as it permits more individuals to survive and reproduce, and also gives the corresponding population more time to evolve higher genome complexity and richer genetic diversity. Nevertheless, except for those with very low allele frequencies that can be driven to extinction by genetic drift, most recessive RLMs are probably retained in diploids or polyploids forever. Given enough time, their population-level allele frequencies could increase to levels that render a diploid (or even polyploid!) organism functionally haploid (Birdsell and Wills, 2003). At this point, any additional mutation in the still functional sister alleles would lead to the death of the individual. In real world populations, such fatality could be rare initially, but rises gradually as the allele frequencies of more and more recessive RLMs trend higher. Eventually a critical point could be reached when the death and/or sterility becomes so common that the critical population size required for species persistence can no longer be assured.

Organisms with Haploid Genomes Purge Obligate RLMs Real-Time

Unlike diploids or polyploids, species with haploid genomes lack a back-up set of sister alleles for genes incurring RLMs. As a result, obligate RLMs are always dominant at their emergence,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

leading to their immediate purging from the descendant population. Put differently, organisms with haploid genomes, with bacteria among the best-known examples, expunge obligate RLMs as they emerge, thus retaining few RLMs in essential genes. Admittedly, these relatively simple organisms could be prone to environmental hazards (e.g. UV irradiation) that increase the rate of mutation, hence RLM emergence, and/or compel the expression of conditional RLMs. Nevertheless, their prolific reproduction under favorable conditions could permit rapid proliferation of individuals that survived the preceding environmental onslaught, thus also the adaptive traits responsible for such survival, perhaps offsetting frequent encounters with hazardous environments.

Among the most notable adaptive traits proliferated in diverse bacterial species is the capacity to internalize DNA molecules present in their surroundings, also known as competence (Birdsell and Wills, 2003 and references therein). Such DNA molecules may have been released by sister bacteria succumbed to environmental assaults. It is further conceivable that some of the internalized DNA molecules could be utilized to complement conditional RLMs in the recipient genomes. Such horizontal gene transfer events were probably the earliest form of sexual hybridization. This form of primitive sex would have provided a selective advantage to bacterial individuals encoding the DNA-internalization capacity, under conditions that were sufficiently extreme to expose the lethality of numerous conditional RLMs simultaneously, causing bacterial lineages lacking this capacity to perish. Thus, it is possible that sexuality first evolved as a back-up mechanism to ensure the survival of haploids from concurrent expression of large numbers of conditional RLMs.

Conditional RLMs Abolishing Community-Level Cooperation in Haploids Are not Efficiently Purged Unless the Cognate Population is Reproductively Bottlenecked

Single-celled haploids like bacteria reproduce through binary fission of parental individuals. Nevertheless, bacteria are also known to encode genes that allow them to cooperate among large numbers of conspecific individuals under certain conditions, especially at sites of natural infections (Aguilar et al., 2007; Henke and Bassler, 2004; Mukherjee and Bassler, 2019; Svenningsen et al., 2009). Examples of such cooperations include crown galls in grapevine resulting from cooperative efforts of millions of *Agrobacterium tumefaciens* individuals, and root nodules in soybean involving numerous individuals of the bacterium *Bradyrhizobium japonicum* (Nakei et al., 2022; Zhu et al., 2000). Biofilms formed by human pathogenic bacteria similarly entail cooperations of high numbers of bacterial individuals (Nadell et al., 2016; Xavier and Foster, 2007). One would expect that bacterial genes responsible for cooperative actions would incur conditional RLMs just like any other genes. It would then be a serious challenge for bacteria to purge these conditional RLMs, or at least minimize their population-level allele frequencies, while engaging in cooperations. This is because such cooperations are commonly made possible by proteins and/or metabolites secreted by nearly all of the conspecific bacteria into the shared community. The well-characterized process of crown gall induction involves hundreds of proteins, opines, and other metabolites that are coordinatively produced by numerous *A. tumefaciens* individuals (Zhu et al., 2000). As a result, a rare *A. tumefaciens* individual with a conditional RLM preventing it from producing one of the proteins/opines/metabolites could survive by accessing the shared goods provided by others, and may even hold a competitive advantage as it no longer diverts energy to the lost functionality (e.g. to translate proteins from the loss-of-function genes). On the other hand, if such conditional RLMs are not purged promptly, eventually all bacterial individuals could incur mutations in one or more genes responsible for various steps of the cooperation. These bacteria may be able to survive in a special niche where they manage to complement one another, but have little chance of surviving at a subsequent infection site as it is unlikely that the complete suite of complementing individuals will all translocate to the new site as a precisely programmed bundle.

What might be the strategy *A. tumefaciens* evolved to address this challenge? It is well known that the vast majority of *A. tumefaciens* individuals that participate in the construction of crown galls become trapped inside the structure, and have no chance of passing their genes, including the conditional RLMs therein, to future generations of *A. tumefaciens* individuals through reproduction. Thus, aside from providing the protective environment for agrobacterial survival and reproduction, the crown gall is also the very instrument that ensures extremely few agrobacterium individuals have the chance to parent descendants. Such reproductive population bottlenecking can be expected to be evolutionarily selected because it isolates the reproducing bacterial individuals from each other in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

separate galls, whose reproduction then produces clonal lineages that originate from single founders in the recent past. Should a founder bacterium harbor a conditional RLM, the same RLM would be inherited by all of its descendants. These descendants, being amplified in isolation, are more likely to translocate to a new infection site together. The conditional RLM carried in their genomes, now lacking access to a complementing sister allele, is then forced to express its lethal phenotype and thereby exterminate the RLM itself. This is to say that they would all fail to perform one specific collaborative function essential for their own reproduction, thereby causing the RLM they carry to be expunged from the population. In short, reproductive population bottlenecking in these bacteria serves the purpose of efficiently purging conditional RLMs in genes encoding diffusible products sharable by many conspecific bacteria co-residing the same reproduction site.

The critical importance of bottlenecked reproduction is not limited to *A. tumefaciens*, as cooperations among large numbers of bacteria are extremely common. Another example concerns a large collection of bacterial pathogens of humans (e.g. *Yersinia pestis*, *Salmonella enteria*) as well as plants (e.g. *Pseudomonas syringae*) that use Type III secretion systems (T3SSs) to deliver up to hundreds of effector proteins to host cells in order to overcome or evade host defense responses (Chang et al., 2014; Coburn et al., 2007). Effector production by large numbers of conspecific bacteria colonizing a single infection site is believed to be essential for robust counter-defense. As a result, conditional RLMs in the genes encoding individual effectors, as well as those encoding T3SS components, if present in one or a few bacterial individuals, are not effectively purged without being subject to reproductive bottlenecking at infection sites.

The strongest evidence for rigorous reproductive population bottlenecking at sites of bacterial infections was provided by a recent study (Aggarwal et al. 2023). Authors of this study constructed a *Streptococcus pneumoniae* population consisting of 2,764 isogenic variants distinguishable from each other by virtue of unique chromosomal barcodes. They then used this *S. pneumoniae* population to colonize the nasopharynx of infant mice. They found that "within 1 day post inoculation, diversity was reduced >35-fold with expansion of a single clonal lineage." (Aggarwal et al., 2023). Although the samples analyzed were nasal and lung washes rather than individual colonization sites, the overwhelming enrichment of a single variant observed was strikingly similar to earlier observations documenting enrichment of single virus variant in plant and animal tissues experiencing systemic virus infections. This leaves little doubt that the barcoded population was subject to extremely tight bottlenecking that limited the reproducing individuals to as few as one per colonizing site. Importantly, careful examination by authors demonstrated that "neither clonal abundance, genetic drift, nor in vivo adaptation explains the observed loss in diversity"; and "host factors and co-infection do not contribute to loss of diversity during colonization" (Aggarwal et al., 2023).

One might argue that despite being introduced as a pre-mixed population, it is possible that a single bacterium from the population launched infection by chance. However, this idea was rejected by one key piece of evidence in the paper. The authors found that loss-of-function mutation within *blpC*, one of the peptide pheromones involved in quorum sensing, caused a partial relaxation of the bottlenecking, permitting a 5-fold increase in clonal diversity. This and other results suggest that (i) multi-bacteria reproduction at the same infection site is potentially possible; and (ii) failure to materialize such multi-bacteria reproduction was due to bacterium-encoded genetic mechanisms, among them quorum sensing, that ensure the number of bacteria permitted to reproduce at each infection site is bottlenecked to as few as one. It should be further noted that quorum sensing is itself a community behavior entailing multiple genes that cooperate among themselves to produce, secrete, and perceive extracellular signals, thus must be reproductively bottlenecked to allow for natural selection-based preservation of its functionality.

How Did Cells with Haploid Genomes First Arise?

Reproductive population bottlenecking discussed above limits the number of reproducing bacterial individuals to very few at each infection site. Given that such bottlenecking enables conspecific bacteria colonizing the same infection site to purge conditional RLMs encoding intercellularly sharable products, how do organisms expunge RLMs in genes encoding products potentially sharable by multiple genome copies occupying the same cell? This question might sound hypothetical because most cells we know have either haploid or diploid genomes. A few known polyploids (e.g. hexaploid wheat) really behave as diploids at reproduction, and the allelic genes

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

encoded on the multiple pairs of genomes, though highly similar, are not identical (Sertse et al., 2023). Nevertheless, multi-nucleus cells do occur either as a result of pathogen (virus, nematode)-induced syncytia, or endoreplication (de Almeida Engler and Gheysen, 2013; Jessie and Dobrovolny, 2021). More recently, Volland and colleagues (Volland et al., 2022) discovered a single-celled, centimeter-long giant bacterium that harbored more than a half million copies of a very large genome. In light of these observations, it is worth asking what evolutionary force(s) are at work to prohibit most cells, especially those destined for reproduction, from housing multiple copies of the same genome. The answer to this question is also tightly connected to another question: how did reproductive population bottlenecking first emerge?

It is now generally accepted that life first emerged as replicating RNAs that were both carriers of inheritance information and replicases (ribozymes) that catalyzed their own replication. The replication of such RNA replicons is thought to have taken place in certain partially enclosed microenvironments supporting appropriate combinations of simple organic compounds (e.g. nucleotides), minerals, solvents (e.g. water), temperature, and light (Joyce and Szostak, 2018; Koonin and Martin, 2005) that were conducive to RNA polymerization. Examples of such microenvironments include the microscopic pores inside rocks arisen through volcanic eruption, or hydrothermal vents formed by underwater volcanic activities (Jerome et al., 2022; Russell, 2007). However, it should be noted that such inorganic microenvironments, unlike modern cells, were unlikely to have co-evolved with the RNA replicons they housed. This is because neither their physical structures, nor contents, were genetically coupled to the replicon RNA synthesized therein.

Furthermore, it is just as important to note that at the dawn of their emergence, multiple self-replicating RNA replicons, as well as their descendants, could have co-existed and co-reproduced in shared microenvironments. This is because both the inheritance carrier and replication catalyzer functions were executed by the replicon RNA themselves without releasing any diffusible (hence sharable) products. As a result, multiple copies of a replicon could co-exist in the same environment without jeopardizing the independent evolution of each other. This in turn should have led to two natural selection inevitabilities. First, the self-sufficiency of these co-existing RNA replicons would have guaranteed that the evolutionary outcomes exclusively fed back to their respective descendants. Meanwhile, sharing the same microenvironment would have facilitated competition among the descendants, leading to dominance by the ones that replicated faster, with higher fidelity, and/or were less prone to destruction.

However, conflicts probably arose not too long after such replicons came into being. The best understood conflict was between replication processivity demanding less structured RNA, and the replicase activity demanding more structured RNA (Joyce and Szostak, 2018). Any evolutionary innovation that resolved this conflict would probably give a selective advantage to the cognate RNA replicon. One such innovation could be for the RNA replicon to evolve the ability to produce a stand-alone replicase ribozyme. This innovation, by separating the execution of replicase function from the replicon itself, could theoretically allow the replicon genome and the replicase ribozyme to assume different structures to suit their respective roles.

Though holding the promise of improved replication efficiency, this evolutionary innovation would have been immediately vulnerable to exploitation. The stand-alone replicase ribozyme, now physically separable from the RNA replicon encoding it, could have just as likely been recruited by highly homologous sister replicons lacking the coding capacity, thus bolstering their own replication. Keep in mind that such innovation, most likely made possible by mutations in the replicon RNA (genome), initially would have occurred in very few replicon copies, if not just one. Therefore, in a shared environment, it was impossible to guarantee that the diffusible replicase exclusively benefited the very replicon encoding the replicase, or its descendants. Put differently, the replicon encoding this presumably advantageous trait could not give rise to more descendants than its sister replicons lacking this coding capacity, thus had little chance to commence the virtuous cycle of natural selection in the shared environments.

Worse, some of the sister replicons, instead of emulating this beneficial innovation, could have evolved a contrarian strategy that enabled them to more efficiently utilize the stand-alone ribozyme, propelling preferential amplification of their own descendants. Admittedly, such preferential dominance of cheaters would be short-lived because such dominance would by itself dilute out replicons encoding the new ribozyme. However, similar cheaters could arise repeatedly, and their

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

short-lived dominance-demise cycles would have decimated the replicons supplying the new ribozyme, leaving the latter little chance to dominate the shared environment.

In theory, many other innovations could have evolved that resulted in potentially beneficial diffusible products. For example, instead of relying on activated nucleotides (NTPs) randomly generated by physical forces in the surroundings, RNA replicons could have evolved the ability to activate nucleotides (converting NMP or NDP to NTP) themselves. However, this trait would have likewise encountered exploitation as the activated NTPs could have been accessed by sister replicons as well. A further example would have been to encode simple peptides with the potential to augment replication and/or stabilize the replicon (Müller et al., 2022), which would have faced the same difficulty of feeding the benefit back to the descendant replicons encoding this very capacity with meaningful exclusiveness. Indeed, we could anticipate only one kind of trans-acting products that might preferentially benefit the cognate replicons. These would be ribozyme activities that specifically destroy sister replicons lacking such coding capacities, leading to exclusive enrichment of replicons encoding them. However, such an activity could be easily defeated by sister replicons evolving to evade the fratricidal ribozymes. The take-home message from the above thought experiment is that the capacity of encoding any diffusible products with potential benefits, emerging in a few isolated mutant replicons, could not have evolved in an environment shared by many homologous RNA replicons lacking such capacities.

By contrast, such evolution would have been possible if the replicons first evolved the capacity to erect some form of active compartmentalization. While lipid-based enclosures have been speculated as the most probable form of primitive cells (Joyce and Szostak, 2018; Koonin and Martin, 2005), it is not unthinkable if the earliest compartments were made up of replicon RNAs themselves, or with them as scaffolds that anchored the deposition of other compounds (Huang et al., 2020; Li et al., 2021).

Another possibility would be for replicon RNAs to acquire ribozyme activities that catalyzed the aggregation or polymerization of simple organics to form polymers that in turn became building blocks of the compartments, such as fatty acids or isoprenoids. Compartments formed this way might not have replicon RNAs as scaffolds, but were nevertheless genetically guided by the replicons. This is akin to bird nests. Though mostly free of body parts of the birds, bird nests are nevertheless bird-specific in terms their shapes, sizes, locations, and compositions, hence must be genetically programmed by the cognate birds – a form of extended phenotype (Dawkins, 1982). Importantly, such a ribozyme activity could be executed by replicon RNAs assuming alternative structures, thus did not necessitate stand-alone entities separable from the replicons themselves. Regardless of shapes or compositions, these earliest, albeit genetically guided compartments would have enabled at least partial separation of replicons with distinct evolutionary characteristics, allowing these characteristics to feed their benefits/defects back to the very replicons encoding them.

Note that while such compartments would have been of particular importance for the evolution of coding capacities leading to products that function in trans, they could have conceivably evolved earlier to partition replicon RNAs with distinct cis-acting properties. For example, a replicon could have evolved to remain stable for a longer time, but replicated slowly. The enhanced stability trait could not evolve if a sister replicon replicated more rapidly to over-populate the shared environment. However, in compartmentalized environments the two replicons could have evolved free of incessant competition against each other, giving both replicons the chance to stabilize their respective traits. As a result, compartmentalized replication could allow one starter replicon to evolve into multiple descendant lineages with distinct characteristics, permitting the survival of at least some lineages in uncertain environments.

The earliest compartments needed not to be perfect – they just needed to facilitate modest levels of differential enrichment for different replicons in separate compartments. For example, a mutant replicon encoding a beneficial diffusible product might not have stood a chance if co-existing with sister replicons at a 1-to-100 ratio, but would have been much more likely to gain dominance if isolated in compartments containing sister replicons at a 1:10 ratio. Indeed in vitro evolution experiments showed that periodical dilution, which served similar purposes as partial compartmentalization, was necessary and sufficient for cell-free evolution systems to stabilize novel traits (Mizuuchi et al., 2022; Okauchi and Ichihashi, 2021).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

It is important to acknowledge that compartmentalization as postulated above is not the same as bottlenecking. This is because all replicon RNA copies sharing the same compartment, including their descendants, could in theory replicate without being bottlenecked. Nevertheless, bottlenecking could be expected to evolve in such compartments. This is because, as new compartments arose through either fission of existing compartments, or de novo assembly, the group of replicon RNAs entering the same new compartment may or may not have descended from the same parental replicon. As a result, a group that did comprise descendants of the same parental replicon would have a comparative advantage, as these descendants would all manufacture the same trans-acting product, feeding the corresponding benefit back to the next generation progeny that also inherited this coding capacity, with the maximal exclusiveness. The relative advantage conferred by a homogeneous replicon group would in turn have selected for intra-compartment bottlenecking that limited the number of replicating replicons to just a few, possibly as few as one, per compartment, thus the emergence of reproductive population bottlenecking.

Furthermore, it would have been advantageous for replicon RNAs excluded from replication by bottlenecking, which would be most replicons in the group, to evolve specialized supporting functions that enhanced the replication of their bottleneck-escaping peers. This is because the entire group was co-ancestral, thus the replicating copies would most likely have encoded the specialized, replicating-enhancing functions as well, and hence passing such coding capacities to their descendants. Examples of such specialized supporting functions could include enlisting most replicon RNA copies as scaffolds of the compartments, or endowing them with ribozyme activities catalyzing the synthesis of fatty acids or isoprenoids. Bear in mind the replicon copies dedicated to such functionalities needed not to be genetically different from their replicating siblings. An easy-to-understand metaphor would be worker ants in an ant colony. Though all genetically capable of becoming a queen (more later), these worker ants nevertheless assume the role of serving a single queen. To reiterate, such functional specialization could evolve as replication of a few at high efficiency could have compensated for the non-replication of the rest that assumed specialized non-replicating roles. In short, reproductive bottlenecking and functional specialization could have evolved hand-in-hand, laying the foundation for the emergence of primitive cells.

Once emerged, further evolution of bottlenecked compartmentalization must have been mutually accommodative with the evolution of more sophisticated coding capacities, trending towards more stringent bottleneck sizes and more diverse, diffusible gene products, eventually leading to the arrangement of one genome per cell – thus the birth of haploid cells. They could also have created the conditions for the evolution of the earliest viruses, as some replicon RNAs in the same compartment could stumble upon opposite evolutionary trajectories. Specifically, instead of evolving additional, supportive coding capacities that together enforce genome bottlenecking through more elaborate cells, some of the replicon RNAs could become primitive viruses by evolving the ability to power more frequent escapes from the existing bottlenecks, hence replicating themselves to outnumber the sister replicons. However, such short-term replicational bursts would inevitably meet the fate of error catastrophe unless they evolved to encode new bottlenecking strategies that specifically safeguard their own population (more later). One could envision that such parasitic lifestyle could survive under two circumstances. One, its reproduction became constrained by counter evolution from the hosting compartments to levels that favor peaceful co-existence. Two, the parasitic replicon acquired the ability to traffic between different host compartments, rendering the survival of host compartments less relevant.

In summary, the haploid cells with one genome copy per cell probably have arisen from positive selection favoring genome compartmentalization that allowed natural selection to act on genes encoding intracellularly diffusible products. Such products, including proteins, metabolites, and simple organic compounds, could be shared by multiple genome copies co-residing the same cell, hence were prone to exploitation. Haploid cells and their predecessors, by reproductively isolating/bottlenecking different genome/replicon copies in separate compartments, minimize such sharing, phenotype masking, or even cheating. They are hence themselves evolutionarily selected as they ensure prompt proliferation of beneficial mutations and swift purging of lethal ones. In short, haploid cells were probably evolved as a form of bottlenecking that, at the dawn of cellular life, ensured the evolution of trans-acting gene products from RNA replicons.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

Finally, we wish to stress that bottlenecking of reproductive populations by itself is stochastic and indiscriminatory. Nevertheless, in a population (e.g. the RNA replicons described above) where beneficial traits emerged in a minute fraction of a large population, bottlenecking provided such rare traits a disproportionate chance to express their phenotypes and be selected. By contrast, in a population where lethal mutations (aka RLMs) are low-frequency events, bottlenecking serves the opposite role of forcing lethality onto individuals carrying such mutations. Yet another easy-to-overlook role of reproductive population bottlenecking is that it blocks most RLMs from expressing the encoded phenotypes by simply denying an overwhelming majority of genome copies/cells, with or without RLMs, the chance to reproduce. Of course, this arrangement also blocks most of potentially beneficial mutations from having descendants. However, the promise of future benefit, being uncertain in nature, cannot be expected to outcompete the need to escape from the certainty of death.

Purging RLMs in Viruses through Intracellular Bottlenecking of Reproducing Viral Genomes

If compartmentalization and bottlenecking are so important both for purging RLMs and for fixing beneficial mutations, why do viruses occupy each infected cell with thousands of genome copies? To address this question, we must differentiate among several possible causes for multi-virus co-existence in the same cell. (i) it could be caused by many copies of the same virus entering the same cell collectively. (ii) it could be caused by rapid and efficient reproduction of a single viral genome copy that overcame the cell boundary. (iii) it could be caused by many copies of a virus entering the same cell collectively, and all of them becoming replication templates. (iv) some or all descendant genomes could have repeated replication for multiple rounds in the same cell. (v) many copies of a virus that have entered the same cell together are subject to intracellular population bottlenecking, permitting just one or very few copies to launch highly efficient replication to produce large amounts of descendants.

Research during recent years provided strong support for the last scenario. Specifically, plant cells loaded with thousands of viral genome copies were found to support the replication of no more than 7 copies, sometimes as few as one (Miyashita et al., 2015; Ren et al., 2023; Zhang et al., 2017). Further mechanistic investigations revealed that at least for one virus [turnip crinkle virus (TCV) with a single-stranded, (+)-strand RNA genome], the intracellular bottlenecking was furnished by a virus-encoded protein which we call bottleneck-enforcing protein or BNEP (Perdoncini Carvalho et al., 2022; Qu et al., 2020). BNEPs likely invoke a multimeric state that self-perpetuates by coalescing additional BNEP monomers and possibly viral genomes they bound to. This Bottleneck, Isolate, Amplify, Select (BIAS) model provides a compelling mechanistic framework that explains both purifying selection against lethal viral mutations, and positive selection for beneficial mutations. We wish to highlight that at least for (+)-strand RNA viruses like TCV, the thousands of genome copies excluded from replication by bottlenecks are themselves contributors to the bottleneck erection, as they collectively serve as mRNA to translate BNEPs to the high concentration needed for the BNEP multimerization.

Consistent with recent evidence suggesting that emergence of viruses predated the last universal cellular ancestor (LUCA) (Krupovic et al., 2020), it is probably reasonable to speculate that viruses descended from some primordial RNA replicons that took a parasitic turn, yet co-evolved with their sister replicons that acquired cell-based compartmentalization (see earlier). Nevertheless, the valuable hint from the virus examples is that cooperation among multiple copies of the same replicon was probably essential for the success of a few from the very beginning of life.

Many Multicellular Organisms Decelerate RLM Accumulation by Alternating between Asexual and Sexual Reproduction

As discussed earlier, conditional RLMs can emerge, accumulate, and spread in a conspecific population for multiple generations in a relatively constant environment. Conversely, sudden environmental upheavals may force concurrent expression of large numbers of conditional RLMs, condemning the population to massive mortalities. Under such circumstances, primitive sexual reproduction in the form of horizontal gene transfer would be a positively selected trait as it ensures the survival of higher numbers of single-celled haploids by using DNA acquired from extracellular environments to complement or repair conditional RLMs. It is perhaps unsurprising that this role of sexual reproduction is preserved in multicellular haploids. Intriguingly, sexual reproduction in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

diploids also serves the opposite role of constraining RLM proliferation. Specifically, it decreases the allele frequency of an existing recessive RLM by (i), creating descendants homozygous for this recessive RLM, thereby expunging it from the population; and (ii), regenerating descendants homozygous for the wildtype allele of this recessive RLM, thus freeing them from this RLM.

The model bryophyte *Marchantia polymorpha* is a vivid example of a multicellular organism with a predominantly haploid genome that exercises asexual reproduction most of time (Shimamura, 2016). It is a tiny plant with a body (thallus) size of no more than 10 centimeters (cm). The *M. polymorpha* bodies recognizable by human eyes are mostly haploid individuals. Although these haploid individuals are of opposite sexes (either male or female), most of the time they reproduce asexually by releasing clonal progeny known as gemmae (Kato et al., 2020; Shimamura, 2016). Nevertheless, this asexual reproduction process is bottlenecked, occurring only in specialized structures called gemma cups. Under favorable environmental conditions asexual reproduction of *M. polymorpha* is very efficient, with every thallus hosting multiple gemma cups, and each gemma cup releasing hundreds of gemmae. These gemmae in turn develop into new thalli that repeat the bottlenecked, asexual reproduction. Such repetition can persist for multiple cycles under favorable conditions (Kato et al., 2020; Shimamura, 2016; Yasui et al., 2019). Thus, *M. polymorpha* is predominantly a multicellular haploid that reproduces through an asexual route. It only undergoes one cycle of sexual reproduction when the surrounding environment becomes hostile, with the resulting diploid stage (sporophyte) almost immediately undergoing meiosis to produce haploid spores from which haploid individuals develop (Kato et al., 2020; Shimamura, 2016).

Given that *M. polymorpha* spends most of time undergoing repetitive asexual reproduction of a haploid genome, it would be safe to assume that obligate RLMs are very efficiently purged as soon as they emerge. Nevertheless, conditional RLMs still emerge and accumulate through multiple cycles of asexual reproduction under permissive conditions. As a result, sudden deterioration of environmental conditions could doom an *M. polymorpha* population through at least three mechanisms. First, it could compel the expression of conditional RLMs. Second, it could cause the genome replication process to be more error-prone, thereby introducing more RLMs. Thirdly, it could greatly reduce the rate of reproduction, rendering the consequence of RLMs disproportionately dire. At these moments, individuals that encode the capacity to undergo one round of sexual reproduction hold a huge advantage because having two sets of chromosomes prevent most RLMs from immediately expressing their lethal phenotypes. Additionally, chromosomal crossovers during sexual mating provide the opportunity to regenerate chromosome copies that are free of RLMs, ensuring the survival of at least some individuals.

We next move to consider multi-cellular diploids, with flowering plants as convenient examples. It is easy to recognize that perennials such as bamboos, aspen trees, and irises routinely undergo bottlenecked asexual reproduction as they produce new shoots from single cells in the roots every spring. Less recognized is the fact that individual branches seen on many trees and flowering plants are products of bottlenecked asexual reproduction as well. These branches originate from single meristematic cells in a manner similar to bamboo shoots, except that the branch-sprouting cells are located in the above-ground parts of plants. Moreover, each of these branches frequently features multiple secondary and tertiary branches, again through bottlenecked asexual reproduction. The seed-producing flowers mostly appear on the tips of branches, but can also emerge on the sides of main stem without branching. Therefore, seed produced by a single flowering plant represent a heterogeneous mixture of genome copies that have experienced varying numbers of asexual reproduction cycles, hence varying chances of RLM introduction. It is clear that asexually reproducing diploids, to which most flowering plants belong, will inevitably accumulate recessive RLMs. Such recessive RLMs, if allowed to persist, could ultimately endanger the survival of the underlying population/species.

We hasten to note that reiterated asexual reproduction cycles, like those occurring in flowering plants, do have advantages. In addition to maximizing the number of descendants, one of the obvious consequences is that dominant RLMs emerging in the genome of the single cells initiating the asexual reproduction events, or any recessive RLMs that gain the chance to express lethality through acquisition of a loss-of-function mutations in their sister alleles, would be eliminated from the future population. In the meantime, diploid genomes probably deters relatively frequent deaths inherent of haploids, permitting the evolution of high genome complexity.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

We now contemplate the fate of recessive RLMs that remain shielded from selection by functional sister alleles in diploid organisms, again using flowering plants as examples. After varying numbers of asexual reproduction cycles, flowering plants go through a single round of sexual reproduction to produce seed that can be more easily dispersed afar. Intriguingly, sexual reproduction in many flowering plants takes place in the same flower (self-pollinating), essentially a scrambled rematch of the same set of genes, except for those disrupted or repaired by meiotic crossovers. As a result, a recessive, obligate RLM present in the cell that initiated the latest round of asexual reproduction would have a 25% probability to become homozygous, thus eliminated by purifying selection. Approximately 50% of the descendants, or 66.7% of reproductively viable descendants, continue to carry the RLM in the recessive form. Strikingly, the remaining 25%, or 33.3% of reproductively live descendants, would be homozygous for the wildtype allele, hence are freed from the RLM. Strikingly, this is achieved through the phenotype-independent self-pollination!

More significantly, thanks to the phenotype-independent nature of this process, it is also effective at clearing conditional RLMs in 25% of descendant individuals. Namely, any given recessive, conditional RLM would be cured from ¼ of the progeny descended from self-pollinating sexual reproduction. Viewed from a different angle, every round of self-pollinating sexual reproduction decreases the population-wide allele frequency of any existing recessive, obligate RLM by 1/6 (from 50% to 33.3%). Meanwhile, although such sexual reproduction does not decrease the overall allele frequency of conditional RLMs, it does erase such RLMs from the genomes of 25% descendants. Assuming no emergence of new RLMs in the same gene, the self-pollinating sexual reproduction practiced by flowering plants could cure a recessive, obligate RLM from more than 50% individuals with mere three repetitions, and continues to reduce their population-level allele frequencies through additional repetitions. Therefore, sexual reproduction involving closely related parents is an effective way to curtail allele frequencies of recessive RLMs, and to ensure any given descendant is freed from a substantial subset of RLMs.

Note that diploid individuals have no way of expunging all recessive RLMs. Nevertheless, sexual reproduction, especially among close relatives, has the unique power of giving rise to individuals that each harbor a different set of RLM-free genes. Obviously, the potency of the RLM-surveillance through sexual reproduction would weaken as the relatedness of sexual partners diminishes. However, with *Homo sapiens* as notable exception, most other organisms reproduce sexually by pairing with partners residing in the same or adjacent ecological niches, thus probably sharing a common ancestor in their recent past. Such sexual activities between related individuals should still serve to lower allele frequencies of obligate RLMs, and regenerate individuals that are freed from a subset of RLMs, albeit less efficiently.

It should be noted that we are not the first to become aware of this role of sexual reproduction. Neiman and colleagues (Neiman et al., 2010) observed that in *Potamopyrgus antipodatrum*, a New Zealand snail with mixed sexual/asexual populations, mutations in mitochondrial genomes accumulated much more rapidly in asexual lineages, hence justifying sexual reproduction as an evolutionary adaptation to purge recessive lethal and deleterious mutations. Similar overaccumulation of deleterious mutations have also been found in asexual populations of stick insects (Bast et al., 2018). In a much earlier study, Mortimer and colleagues (Mortimer et al., 1994) observed that among natural isolates of budding yeast (*Saccharomyces cerevisiae*), those with high levels of homozygosity produced higher numbers of viable spores, whereas the more heterozygous ones produced spores that scored low in viability. Furthermore, some heterozygous strains were found to undergo self-mating to produce homozygous diploid clones that were genetically superior. The authors named this process as "genome renewal" (Mortimer et al., 1994).

We must add that sexual reproduction imposes another layer of genome bottlenecking to the underlying organisms, whose stringency varies among different species. However, there is no doubt that most germline cells, especially male gametes, are blocked from mating, thus denied the chance to parent descendants. Bottlenecking at this stage likewise prevents the RLMs carried by the bottlenecked germline cells from spreading to progeny populations.

Evolutionary Inventions by Arthropods to Fine-Tune RLM Mitigation

Finally, deviations from exclusively asexual or sexual reproduction by many arthropods warrant additional discussions as they further bolster the argument that sexual reproduction primarily serves

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

the purpose of constraining RLMs. Individuals of certain insects, such as aphids and ants, may appear to behave as independent entities. Nevertheless, they have relatively small body sizes and limited mobilities, thus are prone to attacks by predators, host defenses, and environmental hazards. Most aphids rely on living plants for nutritional needs, thus must adjust their surviving strategies depending on how easily accessible host plants are. Accordingly, these aphids alternate between asexual and sexual reproductions based on environmental conditions (Ogawa and Miura, 2014). In environments where long photoperiod and warm temperature coincide with abundantly available nutrients, aphid reproduction is predominantly asexual (parthenogenesis), involving solely females with diploid genomes. These female aphids give birth to numerous baby females asexually, with the babies in turn producing their own descendants asexually as well. Such rampant asexual reproduction can proceed for up to 30 cycles, giving rise to huge pools of aphid individuals, ensuring the survival of the genes they carry by sheer numbers.

Nevertheless, these asexual reproduction periods under favorable conditions involve only diploid females, thus shielding recessive RLMs from purifying selection, and permitting their steady accumulation in the genomes of individual aphids. As discussed earlier, such accumulation of RLMs, conditional or obligate, puts the underlying aphid population at risk of extinction upon drastic environmental changes. Periodical sexual reproduction appears to be the solution evolved by aphids to manage such RLMs. As the daily photoperiod gets shorter and nutrients become scarce, asexual reproduction of aphid females now give birth to both males and oviparous (egg-producing) females. These males and oviparous females undergo one round of sexual mating to produce winter-hardy eggs that develop into new females the next season (Ogawa and Miura, 2014). Because both mating partners arise asexually, and probably even from the same mother, the primary function of one-round-per-year sexual reproduction is likely to compel the manifestation of lethality encoded by recessive RLMs, by rendering them homozygous in a substantial subset of sexual progeny. Note that such periodic sexual reproduction not only keeps obligate RLMs at bay. Rather, a substantial proportion of conditional RLMs are likely also exposed, primarily because the one-round-per-year sexual reproduction coincides with shortened photoperiod, colder temperature, and shortage of nutrients, conditions obliging the expression of conditional RLMs. In summary, this mode of cyclical parthenogenesis is strikingly similar to the reproduction mode of flowering plants discussed earlier.

By contrast, most eusocial ants build their habitats in secluded locations that are frequently separated from the source of nutrients, thus must transport food supply to these sites through cooperations among thousands of worker ants. Such arrangements pose two unique challenges. First, cooperations at this massive scale would necessitate the ants involved to devote high numbers of genes to inter-individual communications. Indeed, ants are known to encode sophisticated systems dedicated to pheromone production and sensing. Second, conditional RLMs that corrupt these genes cannot be effectively purged without bottlenecking reproduction to a very small number of ant individuals.

Extremely bottlenecked reproduction is exactly what eusocial ants do (Wilson, 2008). Eusocial ant colonies contain thousands to millions of worker ants but mostly just one queen. Even though all worker ants are genetically capable of reproduction, only the queen engages in actual reproduction. More importantly, all worker ants do is to safeguard the reproductive success of the queen, and all the queen does is to reproduce. Even more fascinating is that most eusocial ants reproduce through a specialized haploid-diploid mode. Specifically, males are haploids and females are diploids. Furthermore, haploid males are produced by females through asexual reproduction, whereas diploid females are produced through sexual reproduction. Most strikingly, all sperms a queen needs for her lifelong reproduction are acquired through one-time mating, with a single male. Thus, an entire colony of ant progeny derives themselves from one mating event of one pair of parents. This is an extreme form of genome bottlenecking, considering ants are themselves multi-cellular, free-living individuals.

We do not know which of the two arrangements – massive-scale inter-individual cooperation or extremely bottlenecked reproduction – evolved first. Noteworthy to us is the fact that male ants are produced by females through asexual reproduction. This in effect divides the diploid genome of females into two equal haploid subsets. As a result, recessive RLMs present in the diploid female genomes are forced to express the encoded lethality in haploid males, thereby excluding the corresponding male ants, and the RLMs they carry, from participating in sexual reproduction.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

Admittedly, such an arrangement is incapable of expunging all RLMs, because the females engaging in sexual reproduction still harbor their own recessive RLMs. Nevertheless, by subjecting one half of the sex partners (the males) to constant quality surveillance, it serves to suppress the allele frequencies of RLMs in ant populations. Such frequent removal of RLMs, coupled with the extreme reproductive population bottlenecking, is likely under strong positive selection in eusocial insects, given the high risk of incurring conditional recessive RLMs in genes responsible for inter-individual communications.

Synopsis

With this essay we put forward the hypothesis that RLMs are the primary evolutionary driver for living beings ranging from viruses to multicellular organisms. The rationale for this hypothesis boils down to the following steps of logical extrapolation. First, for any biological entities whose reproduction is guided by inheritable genetic blueprints, aka genomes, RLMs inevitably arise during the process of genome replication. RLM-containing genome copies can persist in a population for many reproduction cycles if the genes disrupted by RLMs encode diffusible, hence sharable products; or if the RLM-encoded lethality is conditional. Absent of surveillance mechanisms that actively suppress the population-wide spread of RLMs, more and more copies of co-ancestral genomes would incur RLMs, and in more and more genes, eventually reaching a point when a fragile balance is maintained by a mutually complementing set of genome copies, each retaining a different combination of functional essential genes. At this point, a single additional mutation could cause the demise of the entire co-dependent population. It is then easy to imagine that should a surveillance mechanism capable of suppressing RLM accumulation be evolved by one of the genome copies, it would confer enormous selective advantage to the descendant population of this genome copy.

In the preceding sections we described examples and deductive contemplations that together suggest a generalized evolutionary trajectory for such a surveillance mechanism. Specifically, we advocate that bottlenecked reproduction (replication, amplification) in compartmentalized (isolated) environments must have been the first to have evolved; and has since been faithfully maintained and steadily reinforced as biological entities evolved increasing levels of complexity. In more complex organisms, enhancements to bottlenecked reproduction have included sexual reproduction, as well as alternation between haploid and diploid genome arrangements. In short, we postulate that the Bottleneck, Isolate, Amplify, Select (BIAS) principle, proposed by us earlier to explain the evolution of viruses, holds explanative power for the evolution of all living beings, with the suppression of RLM frequencies being the relentless selection pressure for its preservation across all forms of life.

References

Aggarwal, S.D., Lees, J.A., Jacobs, N.T., Bee, G.C.W., Abruzzo, A.R., Weiser, J.N., 2023. BlpC-mediated selfish program leads to rapid loss of Streptococcus pneumoniae clonal diversity during infection. Cell Host Microbe 31, 124-134.e5. https://doi.org/10.1016/j.chom.2022.10.015

Aguilar, C., Vlamakis, H., Losick, R., Kolter, R., 2007. Thinking about Bacillus subtilis as a multicellular organism. Growth Dev. 10, 638–643. https://doi.org/10.1016/j.mib.2007.09.006

Amorim, C.E.G., Gao, Z., Baker, Z., Diesel, J.F., Simons, Y.B., Haque, I.S., Pickrell, J., Przeworski, M., 2017. The population genetics of human disease: The case of recessive, lethal mutations. PLOS Genet. 13, e1006915. https://doi.org/10.1371/journal.pgen.1006915

Arnheim, N., Calabrese, P., 2016. Germline Stem Cell Competition, Mutation Hot Spots, Genetic Disorders, and Older Fathers. Annu. Rev. Genomics Hum. Genet. 17, 219–243. https://doi.org/10.1146/annurev-genom-083115-022656

Avenarius, M.R., Hildebrand, M.S., Zhang, Y., Meyer, N.C., Smith, L.L.H., Kahrizi, K., Najmabadi, H., Smith, R.J.H., 2009. Human Male Infertility Caused by Mutations in the CATSPER1 Channel Protein. Am. J. Hum. Genet. 84, 505–510. https://doi.org/10.1016/j.ajhg.2009.03.004

Bast, J., Parker, D.J., Dumas, Z., Jalvingh, K.M., Tran Van, P., Jaron, K.S., Figuet, E., Brandt, A., Galtier, N., Schwander, T., 2018. Consequences of Asexuality in Natural Populations: Insights from Stick Insects. Mol. Biol. Evol. 35, 1668–1677. https://doi.org/10.1093/molbev/msy058

Bell, A.D., Mello, C.J., Nemesh, J., Brumbaugh, S.A., Wysoker, A., McCarroll, S.A., 2020. Insights into variation in meiosis from 31,228 human sperm genomes. Nature 583, 259–264. https://doi.org/10.1038/s41586-020-2347-0

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

Birdsell, J.A., Wills, C., 2003. The Evolutionary Origin and Maintenance of Sexual Recombination: A Review of Contemporary Models, in: Macintyre, R.J., Clegg, M.T. (Eds.), Evolutionary Biology. Springer US, Boston, MA, pp. 27–138. https://doi.org/10.1007/978-1-4757-5190-1_2

Chang, J.H., Desveaux, D., Creason, A.L., 2014. The ABCs and 123s of Bacterial Secretion Systems in Plant Pathogenesis. Annu. Rev. Phytopathol. 52, 317–345. https://doi.org/10.1146/annurev-phyto-011014-015624

Coburn, Sekirov Inna, Finlay B. Brett, 2007. Type III Secretion Systems and Disease. Clin. Microbiol. Rev. 20, 535–549. https://doi.org/10.1128/cmr.00013-07

Dawkins, R., 1941-, 1982. The Extended Phenotype. W. H. Freeman and Company, Oxford and San Francisco.

de Almeida Engler, J., Gheysen, G., 2013. Nematode-Induced Endoreduplication in Plant Host Cells: Why and How? Mol. Plant-Microbe Interactions® 26, 17–24. https://doi.org/10.1094/MPMI-05-12-0128-CR

Ganai, R.A., Johansson, E., 2016. DNA Replication—A Matter of Fidelity. Mol. Cell 62, 745–755. https://doi.org/10.1016/j.molcel.2016.05.003

Goldmann, J.M., Veltman, J.A., Gilissen, C., 2019. De Novo Mutations Reflect Development and Aging of the Human Germline. Trends Genet. 35, 828–839. https://doi.org/10.1016/j.tig.2019.08.005

Henke, J.M., Bassler, B.L., 2004. Bacterial social engagements. Trends Cell Biol. 14, 648–656. https://doi.org/10.1016/j.tcb.2004.09.012

Huang, N., Fan, X., Zaleta-Rivera, K., Nguyen, T.C., Zhou, J., Luo, Y., Gao, J., Fang, R.H., Yan, Z., Chen, Z.B., Zhang, L., Zhong, S., 2020. Natural display of nuclear-encoded RNA on the cell surface and its impact on cell interaction. Genome Biol. 21, 225. https://doi.org/10.1186/s13059-020-02145-6

Jerome, C.A., Kim, H.-J., Mojzsis, S.J., Benner, S.A., Biondi, E., 2022. Catalytic Synthesis of Polyribonucleic Acid on Prebiotic Rock Glasses. Astrobiology 22, 629–636. https://doi.org/10.1089/ast.2022.0027

Jessie, B., Dobrovolny, H.M., 2021. The role of syncytia during viral infections. J. Theor. Biol. 525, 110749. https://doi.org/10.1016/j.jtbi.2021.110749

Joyce, G.F., Szostak, J.W., 2018. Protocells and RNA Self-Replication. Cold Spring Harb. Perspect. Biol. 10. https://doi.org/10.1101/cshperspect.a034801

Kato, H., Yasui, Y., Ishizaki, K., 2020. Gemma cup and gemma development in Marchantia polymorpha. New Phytol. 228, 459–465. https://doi.org/10.1111/nph.16655

Koonin, E.V., Martin, W., 2005. On the origin of genomes and cells within inorganic compartments. Trends Genet. 21, 647–654. https://doi.org/10.1016/j.tig.2005.09.006

Krupovic, M., Dolja, V.V., Koonin, E.V., 2020. The LUCA and its complex virome. Nat. Rev. Microbiol. 18, 661–670. https://doi.org/10.1038/s41579-020-0408-x

Kunkel, T.A., 2009. Evolving Views of DNA Replication (In)Fidelity. Cold Spring Harb. Symp. Quant. Biol. 74, 91–101. https://doi.org/10.1101/sqb.2009.74.027

Li, H., Zhang, K., Binzel, D.W., Shlyakhtenko, L.S., Lyubchenko, Y.L., Chiu, W., Guo, P., 2021. RNA nanotechnology to build a dodecahedral genome of single-stranded RNA virus. RNA Biol. 18, 2390–2400. https://doi.org/10.1080/15476286.2021.1915620

Lynch, M., 2010. Evolution of the mutation rate. Trends Genet. 26, 345–352. https://doi.org/10.1016/j.tig.2010.05.003

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. https://doi.org/10.1038/nrg.2016.104

Meinke, D.W., 2020. Genome-wide identification of EMBRYO-DEFECTIVE (EMB) genes required for growth and development in Arabidopsis. New Phytol. 226, 306–325. https://doi.org/10.1111/nph.16071

Miyashita, S., Ishibashi, K., Kishino, H., Ishikawa, M., 2015. Viruses Roll the Dice: The Stochastic Behavior of Viral Genome Molecules Accelerates Viral Adaptation at the Cell and Tissue Levels. PLOS Biol. 13, e1002094. https://doi.org/10.1371/journal.pbio.1002094

Mizuuchi, R., Furubayashi, T., Ichihashi, N., 2022. Evolutionary transition from a single RNA replicator to a multiple replicator network. Nat. Commun. 13, 1460. https://doi.org/10.1038/s41467-022-29113-x

Mortimer, R.K., Romano, P., Suzzi, G., Polsinelli, M., 1994. Genome renewal: A new phenomenon revealed from a genetic study of 43 strains of Saccharomyces cerevisiae derived from natural fermentation of grape musts. Yeast 10, 1543–1552. https://doi.org/10.1002/yea.320101203

Mukherjee, S., Bassler, B.L., 2019. Bacterial quorum sensing in complex and dynamically changing environments. Nat. Rev. Microbiol. 17, 371–382. https://doi.org/10.1038/s41579-019-0186-5

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

Müller, F., Escobar, L., Xu, F., Węgrzyn, E., Nainytė, M., Amatov, T., Chan, C., Pichler, A., Carell, T., 2022. A prebiotically plausible scenario of an RNA–peptide world. Nature 605, 279–284. https://doi.org/10.1038/s41586-022-04676-3

Nadell, C.D., Drescher, K., Foster, K.R., 2016. Spatial structure, cooperation and competition in biofilms. Nat. Rev. Microbiol. 14, 589–600. https://doi.org/10.1038/nrmicro.2016.84

Nakei, M.D., Venkataramana, P.B., Ndakidemi, P.A., 2022. Soybean-Nodulating Rhizobia: Ecology, Characterization, Diversity, and Growth Promoting Functions. Front. Sustain. Food Syst. 6. https://doi.org/10.3389/fsufs.2022.824444

Neiman, M., Hehman, G., Miller, J.T., Logsdon, J.M., Jr, Taylor, D.R., 2010. Accelerated Mutation Accumulation in Asexual Lineages of a Freshwater Snail. Mol. Biol. Evol. 27, 954–963. https://doi.org/10.1093/molbev/msp300

Ogawa, K., Miura, T., 2014. Aphid polyphenisms: trans-generational developmental regulation through viviparity. Front. Physiol. 5.

Okauchi, H., Ichihashi, N., 2021. Continuous Cell-Free Replication and Evolution of Artificial Genomic DNA in a Compartmentalized Gene Expression System. ACS Synth. Biol. 10, 3507–3517. https://doi.org/10.1021/acssynbio.1c00430

O'Rourke, S.M., Carter, C., Carter, L., Christensen, S.N., Jones, M.P., Nash, B., Price, M.H., Turnbull, D.W., Garner, A.R., Hamill, D.R., Osterberg, V.R., Lyczak, R., Madison, E.E., Nguyen, M.H., Sandberg, N.A., Sedghi, N., Willis, J.H., Yochem, J., Johnson, E.A., Bowerman, B., 2011. A Survey of New Temperature-Sensitive, Embryonic-Lethal Mutations in C. elegans: 24 Alleles of Thirteen Genes. PLOS ONE 6, e16644. https://doi.org/10.1371/journal.pone.0016644

Perdoncini Carvalho, C., Ren, R., Han, J., Qu, F., 2022. Natural Selection, Intracellular Bottlenecks of Virus Populations, and Viral Superinfection Exclusion. Annu. Rev. Virol. 9, 4.1-4.17. https://doi.org/10.1146/annurev-virology-100520-114758

Qu, F., Zheng, L., Zhang, S., Sun, R., Slot, J., Miyashita, S., 2020. Bottleneck, Isolate, Amplify, Select (BIAS) as a mechanistic framework for intracellular population dynamics of positive-sense RNA viruses. Virus Evol. 6, veaa086. https://doi.org/10.1093/ve/veaa086

Redondo, M.J., Steck, A.K., Pugliese, A., 2018. Genetics of type 1 diabetes. Pediatr. Diabetes 19, 346–353. https://doi.org/10.1111/pedi.12597

Ren, R., Zheng, L., Han, J., Perdoncini Carvalho, C., Miyashita, S., Zhang, D., Qu, F., 2023. Intracellular bottlenecking permits no more than three tomato yellow leaf curl virus genomes to initiate replication in a single cell. PLOS Pathog. 19, e1011365. https://doi.org/10.1371/journal.ppat.1011365

Russell, M.J., 2007. The Alkaline Solution to the Emergence of Life: Energy, Entropy and Early Evolution. Acta Biotheor. 55, 133–179. https://doi.org/10.1007/s10441-007-9018-5

Sertse, D., You, F.M., Klymiuk, V., Haile, J.K., N'Diaye, A., Pozniak, C.J., Cloutier, S., Kagale, S., 2023. Historical Selection, Adaptation Signatures, and Ambiguity of Introgressions in Wheat. Int. J. Mol. Sci. 24. https://doi.org/10.3390/ijms24098390

Shimamura, M., 2016. Marchantia polymorpha : Taxonomy, Phylogeny and Morphology of a Model System. Plant Cell Physiol. 57, 230–256. https://doi.org/10.1093/pcp/pcv192

Stålberg, K., Bodestedt, Å., Lyrenås, S., Axelsson, O., 2006. A narrow pelvic outlet increases the risk for emergency cesarean section. Acta Obstet. Gynecol. Scand. 85, 821–824. https://doi.org/10.1080/00016340600593521

Svenningsen, S.L., Tu, K.C., Bassler, B.L., 2009. Gene dosage compensation calibrates four regulatory RNAs to control Vibrio cholerae quorum sensing. EMBO J. 28, 429–439. https://doi.org/10.1038/emboj.2008.300

Volland, J.-M., Gonzalez-Rizzo, S., Gros, O., Tyml, T., Ivanova, N., Schulz, F., Goudeau, D., Elisabeth, N.H., Nath, N., Udwary, D., Malmstrom, R.R., Guidi-Rontani, C., Bolte-Kluge, S., Davies, K.M., Jean, M.R., Mansot, J.-L., Mouncey, N.J., Angert, E.R., Woyke, T., Date, S.V., 2022. A centimeter-long bacterium with DNA contained in metabolically active, membrane-bound organelles. Science 376, 1453–1458. https://doi.org/10.1126/science.abb3634

Wilson, E.O., 2008. One Giant Leap: How Insects Achieved Altruism and Colonial Life. BioScience 58, 17–25. https://doi.org/10.1641/B580106

Xavier, J.B., Foster, K.R., 2007. Cooperation and conflict in microbial biofilms. Proc. Natl. Acad. Sci. 104, 876–881. https://doi.org/10.1073/pnas.0607651104

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 29 September 2023 doi:10.20944/preprints202309.2051.v1

Xin, X.-F., Kvitko, B., He, S.Y., 2018. Pseudomonas syringae: what it takes to be a pathogen. Nat. Rev. Microbiol. 16, 316–328. https://doi.org/10.1038/nrmicro.2018.17

Yang, X., Breuss, M.W., Xu, X., Antaki, D., James, K.N., Stanley, V., Ball, L.L., George, R.D., Wirth, S.A., Cao, B., Nguyen, A., McEvoy-Venneri, J., Chai, G., Nahas, S., Van Der Kraan, L., Ding, Y., Sebat, J., Gleeson, J.G., 2021. Developmental and temporal characteristics of clonal sperm mosaicism. Cell 184, 4772-4783.e15. https://doi.org/10.1016/j.cell.2021.07.024

Yasui, Y., Tsukamoto, S., Sugaya, T., Nishihama, R., Wang, Q., Kato, H., Yamato, K.T., Fukaki, H., Mimura, T., Kubo, H., Theres, K., Kohchi, T., Ishizaki, K., 2019. GEMMA CUP-ASSOCIATED MYB1, an Ortholog of Axillary Meristem Regulators, Is Essential in Vegetative Reproduction in Marchantia polymorpha. Curr. Biol. 29, 3987-3995.e5. https://doi.org/10.1016/j.cub.2019.10.004

Zhang, X.-F., Sun, R., Guo, Q., Zhang, S., Meulia, T., Halfmann, R., Li, D., Qu, F., 2017. A self-perpetuating repressive state of a viral replication protein blocks superinfection by the same virus. PLOS Pathog. 13, e1006253. https://doi.org/10.1371/journal.ppat.1006253

Zhu, Oger Philippe M., Schrammeijer Barbara, Hooykaas Paul J. J., Farrand Stephen K., Winans Stephen C., 2000. The Bases of Crown Gall Tumorigenesis. J. Bacteriol. 182, 3885–3895. https://doi.org/10.1128/jb.182.14.3885-3895.2000

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.