



HAL
open science

From viral democratic genomes to viral wild bunch of quasispecies

Philippe Colson, Wahiba Bader, Jacques Fantini, Pierre Dudouet, Anthony Levasseur, Pierre Pontarotti, Christian Devaux, Didier Raoult

► **To cite this version:**

Philippe Colson, Wahiba Bader, Jacques Fantini, Pierre Dudouet, Anthony Levasseur, et al.. From viral democratic genomes to viral wild bunch of quasispecies. *Journal of Medical Virology*, 2023, 95 (11), <10.1002/jmv.29209>. <hal-04275773>

HAL Id: hal-04275773

<https://hal.science/hal-04275773v1>

Submitted on 14 Nov 2023



HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

From viral democratic genomes to viral wild bunch of quasispecies

Philippe Colson^{1,2,3}  | Wahiba Bader^{1,2} | Jacques Fantini⁴ | Pierre Dudouet^{1,2} | Anthony Levasseur^{1,2} | Pierre Pontarotti^{1,5} | Christian Devaux^{1,5} | Didier Raoult^{1,2} 

¹IHU Méditerranée Infection, Marseille, France

²Aix-Marseille Université, Institut de Recherche pour le Développement (IRD), Microbes Evolution Phylogeny and Infections (MEPHI), Marseille, France

³Assistance Publique-Hôpitaux de Marseille (AP-HM), Marseille, France

⁴INSERM UMR_S 1072, Aix-Marseille Université, Marseille, France

⁵Department of Biological Sciences, Centre National de la Recherche 16 Scientifique (CNRS)-SNC5039, Marseille, France

Correspondence

Philippe Colson and Didier Raoult, IHU Méditerranée Infection, 19-21 Blvd Jean Moulin, 13005 Marseille, France.
Email: philippe.colson@univ-amu.fr and didier.raoult@gmail.com

Funding information

Agence Nationale de la Recherche

Abstract

The tremendous majority of RNA genomes from pathogenic viruses analyzed and deposited in databases are consensus or “democratic” genomes. They represent the genomes most frequently found in the clinical samples of patients but do not account for the huge genetic diversity of coexisting genomes, which is better described as quasispecies. A viral quasispecies is defined as the dynamic distribution of nonidentical but closely related mutants, variants, recombinant, or reassortant viral genomes. Viral quasispecies have collective behavior and dynamics and are the subject of internal interactions that comprise interference, complementation, or cooperation. In the setting of SARS-CoV-2 infection, intrahost SARS-CoV-2 genetic diversity was recently notably reported for immunocompromised, chronically infected patients, for patients treated with monoclonal antibodies targeting the viral spike protein, and for different body compartments of a single patient. A question that deserves attention is whether such diversity is generated postinfection from a clonal genome in response to selection pressure or is already present at the time of infection as a quasispecies. In the present review, we summarize the data supporting that hosts are infected by a “wild bunch” of viruses rather than by multiple virions sharing the same genome. Each virion in the “wild bunch” may have different virulence and tissue tropisms. As the number of viruses replicated during host infections is huge, a viral quasispecies at any time of infection is wide and is also influenced by host-specific selection pressure after infection, which accounts for the difficulty in deciphering and predicting the appearance of more fit variants and the evolution of epidemics of novel RNA viruses.

KEYWORDS

coronavirus, intrahost diversity, mutant, quasispecies, SARS-CoV-2, variant

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

© 2023 The Authors. *Journal of Medical Virology* published by Wiley Periodicals LLC.

1 | VIRUS SPECIES DEFINITION

Viruses are one of the most difficult biological elements to define, and the creation of names to define them by analogy with previously defined elements very quickly shows its limits. Thus, the notion of a virus as a living or nonliving element has given rise to debates that only make sense semantically, as the very definition of a virus cannot be exhaustive and ignores either giant viruses or “selfish” genes without any real boundaries being established between each of these replicative life forms.¹ Under these conditions, the definition of viral species is an almost impossible challenge. Indeed, the definition of species, as first derived from eukaryote species, theoretically assumes a monophyly that is impossible to truly assert for viruses because of their ease of recombination.² For example, we have recently described a recombinant between the vaccine poliovirus and an enterovirus C in Africa³ and recombinants between different SARS-CoV-2 variants,⁴ and it is not clear how they can be classified considering a monophyletic approach. It should be noted that monophyly should be no longer conceivable for other living entities that are likely to harbor genetic sequences from other kingdoms of life.⁵ The massive genetic work of the 21st century has shown that all living entities are in fact mosaics and that no living entity is monophyletic^{6–8}; the idea of a tree of life with monophyletic entities at its extremities and with a unique ancestor (LUCA, last universal common ancestor) has been definitively ruled out of scientific reality.

Furthermore, thinking about viruses has been biased from the beginning of history. Early work on the toxicity of a plant extract (in this case tobacco) led investigators to assume that its sickness was not due to an organism but to a toxin.^{9,10} Later, the toxic product was identified as tobacco mosaic virus containing RNA and proteins, while infectivity by viral RNA was reported.^{11,12} However, crystallization of the tobacco mosaic virus by WM Stanley (Nobel Prize in Chemistry in 1932) was nonetheless the basis for the long-standing view that viruses are inanimate, nonliving objects.¹³ Moreover, other authors, such as FM Burnet (Nobel Prize in Medicine in 1960), who compared viruses to intracellular bacteria, considered that viruses are merely cellular organisms that had lost their reproductive apparatus; in particular, what were subsequently identified as ribosomes (Claude A, de Duve C, Palade GE, Nobel Prize in Physiology and Medicine in 1974).¹⁴ Thus, from the onset, there was an assumption of a molecule originating from cellular living entities and of reductive evolution of previous cellular entities. In 1957, A. Lwoff proposed a concept of virus that largely opposed viruses to cellular organisms by defining viruses based on negative criteria.¹⁵ Regarding viral species, they were later defined in part on the basis of the existence of isolates representative of those species.¹⁶ A viral isolate could be obtained from any culture medium, including animals, plants, and cells, and the perception of these isolates was that they represented a defined object with regular mutations that eventually allowed for adaptation by natural selection. An isolate was considered clonal by definition, with a single nucleic acid sequence. However, the variability of these viruses appeared to be massive through mutations but also through loss of genes identified by the study of

herpesviruses¹⁷ or poxviruses^{18,19} as well as by that of the giant mimivirus in successive culture passages on its host (*Acanthamoeba* spp.).²⁰ Mimivirus could indeed lose 16% of its genome in allopatric culture, particularly the genes not initially transcribed.²¹ This experiment, imitating Pasteur's experiments on fowl cholera with *Pasteurella multocida*,²² showed that the adaptation mechanism could be linked to gene loss and not to point mutations or “slow changes in progress,” as postulated by Darwin.²³ The first sequences of viral genomes (the very first was that of RNA bacteriophage MS2²⁴) were obtained by manual, time-consuming, and tedious techniques that were thereafter replaced by the Sanger sequencing method,²⁵ which made it possible to obtain entire virus genomes relatively easily.

However, from the outset, the perception that a virus had a defined genome was an essential aspect found in virtually all descriptions of viruses until the recent past. The evolution and sophistication of sequencing methods quickly showed that this was not the case and that the frequency of mutations had been grossly underestimated by creating genome definitions based on the most frequent form of genomes observed after sequencing, what has been called the “consensus” genome²⁶ and what we call “democratic” genomes. These genomes represent the majority of genomes but by no means their diversity, and thus representing an isolate by a single sequence is a basic scientific error. In practice, viruses remain very difficult to define. Even today, viral species are not precisely definable unless we create an artifact that we know will have difficulty standing the test of time given the amplification of virus analyses and the tremendous increase in sequencing data. The intra-host genetic diversity and quasispecies have not been incorporated in the current virus taxonomy, in which the species rank is the only one that is not associated with a secondary rank, as no consensus has been reached regarding the “subspecies” definition.^{16,27–29} Moreover, our current ability to understand the importance of recombination between viruses, and between viruses and their hosts, leads to a much more complex view of the nature of viruses. Given this complexity, the discovery and description of quasispecies obtained by deep sequencing methods is leading to a revolution in thinking about the evolution of viruses.

2 | THE CONCEPT OF VIRAL QUASISPECIES AND THEIR GENERATION, EMERGENCE, AND EVOLUTION

Variation was put forward early as a major trait of viruses on the basis of observations of the mutability and rapid evolution of RNA viruses such as influenza virus, Newcastle disease virus, coxsackie A9 virus, and vesicular stomatitis virus.^{30–36} During the 1970s, the nonhomogeneity of RNA viral populations was highlighted by studies of site-specific mutations conducted on Qbeta RNA bacteriophage.^{37,38} This led to the conclusion that this phage was not a “defined unique structure” but a “weighted average of a large number of individual sequences.” This was linked to the high mutation rate of the phage estimated at 10^{-4}

mutations per nucleotide copied and involved a tolerance of each individual genome to an indeterminate proportion of mutations.³⁷ The genetic diversity of RNA viruses was also deduced from the variety of mutant phenotypes observed during the course of persistent infections *in vitro* with animal viruses such as vesicular stomatitis virus³⁶ and foot-and-mouth disease virus.³⁹ This paradigm was strengthened when knowledge of virus genetic evolution considerably expanded during the 1960s through the 1980s due to technological advances in the field of molecular biology and sequencing, which eventually allowed for obtaining a tremendous amount of viral sequences, including from viral genomes and metagenomes.^{24,40–43} Nevertheless, the essentiality of sequences produced, including those produced by next-generation sequencing during the 21st century, have been stored and analyzed in the form of a consensus/“democratic” sequence. This latter sequence is the succession at each of its nucleotide positions of the majority nucleotide among those detected by sequencing in the sample. Two or more coexisting nucleotides can be possibly indicated by an extended International Union of Pure and Applied Chemistry (IUPAC) nomenclature code, but a single nucleotide (the one majoritary) is mentioned in a very large majority of cases, erasing from the consensus/“democratic” sequence any information about other components of the quasispecies. Hence, such a consensus/“democratic” sequence is a simple but inadequate approach to describe and consider the genetic diversity of viruses in a population of hosts, in a single host, or even in a single host sample.

The concept of quasispecies enables far more comprehensive representation of the diversity of sequences than do consensus/“democratic” sequences. The quasispecies theory was developed during the 1970s by Manfred Eigen, a biophysical chemist, and Peter Schuster, a theoretical chemist, to explain the evolution and self-organization of primitive replicons.^{44,45} This theory was built independently from laboratory experiments conducted on the genetic diversity of RNA viruses, but it was noticed that it fitted quite well with the results of these latter experiments. The theory derives from the species concept and from the molecular theory of evolution. A quasispecies was defined as the complex mutant spectrum (or swarm or cloud) produced in replication systems through regular generation of error copies, and it was first named “comet tail.” It echoes at the genetic level the paradigm that populations are usually not composed of fully identical individuals but of “types” distinct between each other by some criteria.⁴⁶ Thus, amplification by copying from a single (“wild-type”) clone generates mutants, and the mutations produced delineate a sequence space that expands depending on the mutation rate and the selection of mutants, and that reaches a stable steady-state mutant spectrum where for each quasispecies component, the contributions of both mutation and selection balance each other. The quasispecies concept is linked to an “error threshold” rate above which genetic information can no longer be propagated, which corresponds to extinction. There are many alternative terms to quasispecies, including intrahost variation, intrahost

diversity, mixture of mutants, nucleotide degeneracy, or heterospecies.

A viral quasispecies is a dynamic distribution of nonidentical but closely related mutants or recombinant or reassortant viral genomes, as viral quasispecies are also known to be the consequence of recombinations and reassortments.^{47,48} A viral quasispecies undergoes continuous genetic variations and is the subject of random drift, competition, and selection and responds to various selective constraints. The existence of viral quasispecies has been described for many RNA viruses, including bovine viral diarrhea virus, West Nile virus, hepatitis A virus, hepatitis C virus, HIV, and most recently SARS-CoV-2.^{49–54} Spectra of viral mutants, recombinants, and reassortants, rather than individual genomes, are the target of evolutionary events. They are the source of viral adaptability, as they comprise continuously changing dynamic reservoirs of genotypic and phenotypic viral variants with given fitness landscapes.⁵⁵ Such intrahost genomic variability leads to antigenic variability and phenotypic specificities. These phenotypic specificities are related to replicative capacity; transmissibility; host range and tropism; susceptibility to immune responses, vaccines, and antivirals; clinical severity of infections; and sensitivity of molecular and serological diagnostic tests.^{48,55–58} Viral quasispecies may also undergo a drastic reduction in population size due to isolation of one or several individuals from the mutant spectrum through bottlenecks during inter- as well as intra-host transmission.⁵⁹ Moreover, viral quasispecies have collective behavior and are the subject of internal interactions that comprise interference, complementation, or cooperation. For instance, a low-fitness mutant can suppress a high-fitness mutant,⁶⁰ or an attenuated virus can suppress a virulent virus.⁶¹ The viral quasispecies concept encompasses an error threshold that delineates the maximal mutation rate of the dominant sequence at which the quasispecies can be metastable, and above which a drift of the sequence population occurs in the sequence space.^{26,48} Considering the above principles, each viral quasispecies is unique in its composition and differs from those that have preexisted and from those that will exist, whether in the same host or in different hosts and whether in an identical context or in different contexts.⁶² It continuously evolves even if the consensus sequence remains the same.

It is worth noting that for a given and unique sequence, diversity can be represented in the form of a profile (i.e., Profile hidden Markov models [HMM]).⁶³ Profiles are constructed by converting multiple sequence alignments into position-specific scoring systems (PSSM). Profile HMMs are probabilistic models including evolutionary changes in a set of homologous sequences. This principle may be applied during genome assembly by integrating the diversity of reads as genomic profiles. Deep sequencing further increases the need to integrate the sequence variability of reads. Ideally, in the future, the complete genomic sequence will be described in the form of a global genomic profile that would take into account the diversity of each position in a genome.

3 | THE HIV PARADIGM: FROM SANGER TO NEXT-GENERATION SEQUENCING

Sequencing of HIV-1 and HIV-2 RNA and DNA has been a major tool to study and manage HIV infection. To date, more than 1.2 million sequences, including approximately 15 089 genomes, have been deposited in GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>), and HIV genotypic resistance testing is an unavoidable tool for managing HIV treatment,⁶⁴ with several algorithms available worldwide to interpret the mutation patterns of genes encoding targets of antiretroviral therapies (e.g., <https://hivdb.stanford.edu/>, https://rega.kuleuven.be/cev/avd/publications/european_guidelines_page <https://www.iasusa.org/resources/hiv-drug-resistance-mutations/>, <https://hivfrenchresistance.org/>). Sequencing of HIV genes developed in the 1990s with the advent of molecular biology techniques applied to diagnosis in clinical virology, particularly that of plasma viral load.⁶⁵ Use of Sanger sequencing for genotypic evaluation of resistance to antiretrovirals appeared in the mid-1990s.^{66,67} Nonetheless, drug resistance genotyping was not immediately recognized as a tool of interest for the management of patients. But sequencing developed alongside phenotypic tests and then replaced them.^{68,69} Above all, the concept of quasispecies was not initially considered in reading and interpretation of sequencing results.

In fact, HIV-1 genetic diversity in infected patients is tremendous. It was early estimated that approximately 10^9 viruses were produced each day by infected cells, with approximately one mutation per replicative cycle for an approximately 10,000

nucleotide-long genome (i.e., mutation rate is 10^{-4}).⁷⁰⁻⁷² Moreover, HIV-1 quasispecies are compartmentalized, for instance, in peripheral blood mononuclear cells, lymph nodes, the genital sphere, or the central nervous system.⁷³ Interindividual transmission of minority drug-resistant HIV-1 variants was evidenced at the time of primary infection by detection of these minority variants in transmitters or by their decreased prevalence after several weeks in seroconvertors in the absence of antiviral therapy, as explained by their lower fitness than wild-type HIV-1.^{74,75} By sequencing HIV-1 genes, the emergence of mutations associated with HIV-1 resistance to antiretrovirals can be detected early not only in “democratic” sequences but also in minority components of the quasispecies, the detection of which can predict their expansion and clinical significance.^{76,77} This notion that minority HIV-1 variants preexist echoes in virology the existence of preadaptive mutations evidenced during the 1950s with the observation of the clonal occurrence of mutants in bacteria based on replica plating.⁷⁸ Sanger sequencing allows for visualization of some minority quasispecies when reading an electropherogram (Figure 1A).⁷⁶ Nonetheless, Sanger technology only enables the detection of genomes representing at least 10%–20% of the overall virus population of the sample.⁸⁰ With the development of next-generation sequencing techniques, detection of mutations representing less than 10% of quasispecies is now possible, with 5% as the prevalence threshold conventionally used to consider these mutations as significantly involved in clinical resistance to antiretrovirals.⁸¹ A map of next-generation sequencing reads generated from the blood of an HIV-1-infected patient (Figure 1B,C) clearly shows that



FIGURE 1 Schematic of HIV-1 quasispecies composition by Sanger (A) and next-generation sequencing (B, C). (A) The electropherogram is a screenshot taken from the computer of an ABI 3500 instrument (Thermo Fisher Scientific). Arrows mark double peaks that indicate mixed populations of sequences. (B, C) Screenshot of mapping reads generated by Illumina technology using a MiSeq instrument (Illumina) of an HIV-1-infected blood sample, as visualized with the Integrative Genomics Viewer tool.⁷⁹ In (B), sequence reads are shown for the same genome coordinates than the electropherogram. Mixed viral populations can only be seen at two positions by the electropherogram, whereas additional mutations are evidenced by Illumina sequencing reads. In (B, C), each gray rectangle is a read. Colored horizontal lines in reads indicate mutations relatively to the “democratic” sequence for the sample. Only a very small fraction of sequence reads can be shown considering the huge sequencing depth. In (C), the region near 2600 bases corresponds to an area of the genome with a low depth of next-generation sequencing reads. As a matter of fact, some reads exist which could have been mapped to this region, but they do not appear on this limited-in-size screenshot of the mapping visualization software. See box 1.

BOX 1. Potential technical issues in deciphering the composition of viral quasispecies

- Exhaustiveness and accuracy in the characterization of the composition of viral quasispecies:** The full composition of the viral quasispecies is not comprehensively known, particularly for mutants below the threshold of 1%. With current technologies and tools, detection of every single molecule composing the quasispecies is not possible. Indeed, the sensitivity threshold of next-generation sequencing, commonly known as “depth bias,”¹⁴² does not allow for accurate exhaustive detection. The estimated sensitivity of Illumina next-generation sequencing to detect minor viral variants in a mixture of standards was reported to be 97.5% for a minor variant present at 1%.¹⁴³ In another work, the lowest threshold of detection of minor variants was estimated at 0.5%–1% with Illumina next-generation sequencing. In addition, the rate of sequencing errors was estimated to range between 0.5% and 0.1% (medians) with Illumina technology (on MiSeq and Novaseq instruments, respectively)¹⁴⁴ and reaches approximately 10% or more with the Nanopore Technology.¹⁴⁵ Such introduction of artifactual mutations hence limits the detection accuracy of highly minority genuine mutations. A greater coverage of the sequences with reads (so-called sequencing depth) increases both accuracy and sensitivity, but sequencing accuracy decreases as the threshold of detection of variants increases due to biases affecting cDNA synthesis, PCR amplification, library preparation, and sequencing itself. In summary, the diversity of the wild bunch can vary considerably according to the tool used to measure it.
- Obtaining of long sequences by next-generation sequencing:** The availability of the longest NGS reads possible, and beyond of whole virus genome sequence is a valuable goal for the characterization of the composition of viral quasispecies. Indeed, a major goal is to produce reads that each are one of the quasispecies component as a single molecule. Alternatively, obtaining only reads shorter than the whole genome length can allow detecting mutations as being part of the viral quasispecies but leaves unresolved whether these mutations are part from a same viral genome or are present in different combinations on different genomes. For instance, currently, Illumina NGS technology only allows generating short reads, in the order of magnitude of a few hundreds bases in length. In contrast, Oxford Nanopore and PacBio technologies allow generating much longer reads that can cover whole viral genomes. In recently published studies, these technologies have been successfully used to perform sequencing of long-read single molecules in the range of 10–30 kilobases and have been applied to HIV, viral hepatitis, and SARS-CoV-2, which allowed obtaining single molecules covering viral genes and even genomes and studying within host viral epistasis.^{146–150} Nevertheless, until recently, the error rate of these latter NGS technologies was high, around 10%,¹⁴⁵ which hampered a very accurate characterization of the quasispecies components. Improvements of these error rates will improve significantly quasispecies studies.

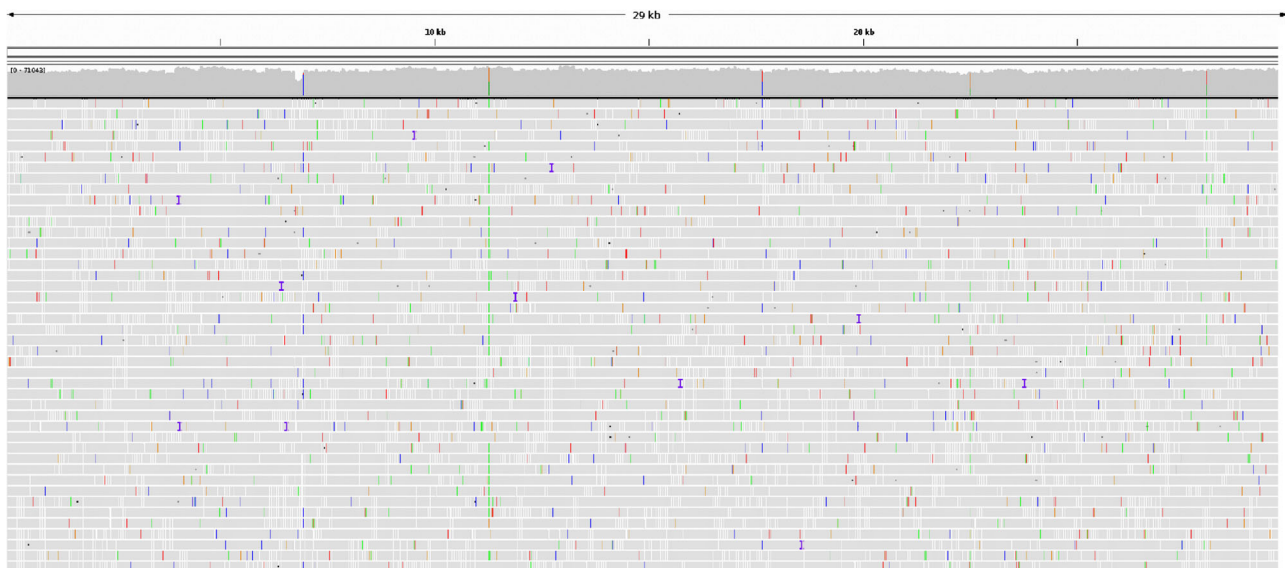


FIGURE 2 Screenshot of mapping of next-generation sequencing reads from a SARS-CoV-2-infected respiratory sample. Reads were generated by Illumina technology using a NovaSeq instrument (Illumina). Mapping was visualized with the Integrative Genomics Viewer tool.⁷⁹ Each gray rectangle is a read. Colored horizontal lines in reads indicate mutations relatively to the “democratic” sequence for the sample. Only a very small fraction of sequence reads can be shown considering the huge sequencing depth. See box 1.

reads are not devoid of mutation(s) compared to the “democratic” genome and that reads are not identical to their mates. In fact, there are non identical sequences, and the “democratic” genome is thus virtual.

4 | REDEFINING VIRUS ISOLATES AS A WILD BUNCH

Our perception of viruses was first that of biomolecules that can be reproduced in a host¹³; eventually, the notion of quasispecies acknowledged the considerable variation of viruses, particularly RNA viruses, which can be thought to play a considerable role in evolutionary strategies.⁴⁸ Until recently, quasispecies seemed to consist of multiple sequences, the most efficient (fit) of which, as represented by the “democratic” genome, can be transmitted and is the only one subject to selective pressures. Another approach is now more reasonable. In fact, it seems that in at least a few cases, it is the whole “bunch” of molecules that is transmitted (Figures 1C and 2). Indeed, the phenomenon of multiple invasive sequences has been well developed for measles virus.^{82,83} This virus has two tropisms: one for lymphocytes and lymph nodes and one for epithelial cells. The most efficient viral form in lymphocytes differs from that in epithelial cells. Several hypotheses have been proposed on this subject. The first is that viral mutants appear during infection, allowing the virus to replicate in one of its targets and to produce a second form from the quasispecies that can reach the second target. Currently, the common hypothesis is that hosts are infected by a whole “wild bunch,” which is distributed according to its tropism in either lymphocytes or epithelial cells. This phenomenon has apparently been confirmed in many situations. In the case of infection of mice with poliovirus, for example, different viral sequences are found in different organs.⁸⁴ The same phenomenon has been reported in SARS-CoV-2 infections, including showing in some cases dominant mutations in certain samples that had not yet been found in the “democratic” genomes identified later in epidemic infections.⁸⁵ This probably means that the conditions for selection of different mutants in different organs differ and allow for different “democratic” genomes to emerge.

In this sense, it is plausible that we can no longer consider virus isolates or strains as being linked to a single sequence but rather as a bunch of sequences. It is possible in these circumstances that the differences in mortality in humans and experimental animals are related to the composition of the bunch, just as it is possible that the role of inoculum size in infection is also related to the extent to which the sequences are dispersed throughout the bunch. This allows for adaptation to different organs, likely resulting in diseases of different natures in terms of virus localization to different organs.

This evolution of thought about viruses was made possible by the evolution of tools, from the Chamberland filter to next-generation sequencing, passing by electron microscopy, which

successively allowed for observation of the toxic effect and observation of the virus, obtaining the first genome by detecting only the majority sequence, discovery of quasispecies and highlighting multiple species as multiple genetic sequences during an infection by “the same virus.” To be established, this latter, new view will require a paradigm shift in the sense of Thomas Kuhn (1962), which will take some time. Indeed, a number of philosophical debates on viruses and even the definition of viruses themselves are still at a prescientific stage, which will make this conceptual approach difficult.

5 | THE CASE FOR COVID-19, INCLUDING SARS-COV-2

In December 2019, a novel coronavirus, named severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), emerged in Wuhan, Hubei Province, China⁸⁶ and later caused a pandemic.⁸⁷ Phylogenetic analyses based on its complete genome revealed that it belongs to the genus Betacoronavirus and subgenus Sarbecovirus, which also encompasses SARS-CoV and Middle East respiratory syndrome (MERS)-CoV, two other coronaviruses causing severe respiratory syndromes in humans.⁸⁸ In addition, the SARS-CoV-2 genome showed high genetic similarity (approximately 96%) with that of the bat coronavirus RaTG13.⁸⁶

SARS-CoV-2 is a linear, nonsegmented, single-stranded, positive-sense RNA virus. Its genome comprises 29 903 nucleotides with 12 open reading frames (ORFs).⁸⁸ As in the case of other RNA viruses, SARS-CoV-2 RNA replication through viral RNA-dependent RNA polymerase (RdRp) is associated with mutations due to the low fidelity of this enzyme. One particularity of coronaviruses is that their genome encodes a 3′-5′ exonuclease (NSP14 in SARS-CoV-2) with corrective activity. Nonetheless, this activity only reduces the SARS-CoV-2 mutation rate, which is estimated to be on the order of 6.7×10^{-4} – 1×10^{-3} /nucleotide/year^{89,90} and was observed to be approximately 9.8×10^{-4} substitutions per site per year,⁹¹ equivalent to approximately 30.4 substitutions/genome/year (<https://nextstrain.org/ncov/gisaid/global/6m?l=clock>).⁹² SARS-CoV-2 mutations may also have other sources and notably occur through the action of apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC), or adenosine deaminases acting on RNA (ADAR) cellular RNA-editing enzymes, which are host antiviral defense mechanisms that can restrict viral infection^{93,94} but also increase the frequency of mutations. Finally, recombinations between SARS-CoV-2 genomes can generate intrahost diversity, and multiple SARS-CoV-2 recombinants have been reported, mostly during the most recent period of the pandemic since 2022.^{4,95,96} Therefore, mutated SARS-CoV-2 genomes are generated at a substantial rate, leading to a soup of nonidentical genomes in which each single genome can differ from another by one or several nucleotides and thus deviates from the “democratic” genome sequence.

5.1 | Quasispecies for non-SARS-CoV-2 coronaviruses

Viral quasispecies have already been reported for coronaviruses other than SARS-CoV-2. In nonhuman animals, the existence of intraindividual quasispecies was reported in 1995 for the mouse hepatitis virus (MHV).⁹⁷ Adami et al. detected an increase in genetic diversity in persistently infected mice, with apparent abnormal accumulation of mutations in the spike gene. Such existence of quasispecies in mice chronically infected with MHV in the spike domain was confirmed in two other studies.^{98,99} Interestingly, in one of these studies, mice with the most severe chronic paralysis were those that harbored the most complex quasispecies.⁹⁹ In 1999, the presence of quasispecies for the feline coronavirus was reported.¹⁰⁰ A study on the bovine coronavirus (BCoV) also observed the presence of spike mutations, with 104 nucleotide changes associated with several intrahost clones within a single sample.¹⁰¹ Another study involving nasopharyngeal samples of pigs with documented HKU15 coronavirus infection detected six quasispecies in a single sample.¹⁰²

Regarding human coronaviruses other than SARS-CoV-2, Viehweger et al. reported in 2019 the presence of diverse subgenomic mRNAs for human coronavirus 229E (HCoV-229E) through direct RNA and long read sequencing.¹⁰³ Additionally, Gorse et al. detected quasispecies for HCoV-OC43 but with a lower intrasample frequency than for hepatitis C virus.¹⁰⁴ Regarding MERS-CoV, Lu et al. reported in 2017 the presence of different spike domain S1 variants (wild type or with a 530 nucleotide-long deletion) in the same human sample.¹⁰⁵

5.2 | Quasispecies for SARS-CoV-2

SARS-CoV-2 was reported in several studies to exhibit genetic intrasample and intrahost diversity in clinical samples and in virus culture. The mutational patterns of the minority quasispecies were heterogeneous according to the studies. One issue is the extent to which the wild bunch is clinically significant, that is, in what proportion minority SARS-CoV-2 variants can be infectious and transmitted between two individuals. It has been estimated that 10^9 to 10^{11} viruses are carried by an infected person during peak infection.¹⁰⁶ Based on a mutation rate of 10^{-3} substitutions per nucleotide per year or 3×10^{-6} mutations per nucleotide per replicative cycle and 3–7 replicative cycles per infection, 10^5 – 10^8 mutations occur per host. Thus, people are infected by a wild bunch and absolutely not by a single, clonal type of molecule.

Studies that reported intrahost SARS-CoV-2 genetic diversity have notably provided data for immunocompromised, chronically infected patients, patients treated with monoclonal antibodies targeting viral spike, and different body compartments from the same patients (Supporting Information: Table S1). Choi et al. described the case of a 45-year-old male patient immunocompromised due to severe antiphospholipid syndrome, with SARS-CoV-2 shedding for 151 days, during which the patient was notably

administered remdesivir and casirivimab/imdevimab monoclonal antibodies.¹⁰⁷ During this period of prolonged infection, a total of 45 nucleotide substitutions, including 24 nonsynonymous mutations, and 34 deletions occurred. In the spike, 12 amino acid substitutions occurred, including the N501Y and/or E484K substitutions and the deletion Y144-. Kemp et al. reported SARS-CoV-2 genetic evolution and reduced sensitivity to neutralizing antibodies in an immunosuppressed male in his 70s who was diagnosed in 2012 with marginal B-cell lymphoma and who was chronically infected with SARS-CoV-2 and notably received convalescent plasma for this infection.¹⁰⁸ Whole-genome sequences obtained by next-generation sequencing from samples collected at 23 time points over 101 days showed substantial shifts in viral quasispecies, notably with emergence of a predominant virus harboring amino acid substitution D796H in the spike protein S2 subunit and a deletion of amino acids 69 and 70 of the S1 N-terminal domain. In vitro tests indicated that the spike mutation D796H appeared to be the main contributor to the reduced susceptibility to neutralizing antibodies and was also associated with reduced infectivity, which might be compensated for by spike deletion Δ H69/ Δ V70. Interestingly, this predominant virus exhibiting an immune escape genotype decreased in prevalence as passively transferred antibodies waned. Laubscher et al. analyzed 12 clinical specimens collected from the upper or lower respiratory tracts of 12 oncological patients and reported intrapatient evolution of minority quasispecies mainly involving the ORF1ab gene.¹⁰⁹ Ip et al. used Illumina next-generation sequencing and nanopore strategies with two respiratory specimens, sputum and saliva collected 2 days apart, from a 75-year-old patient with severe disease.¹¹⁰ They identified the spike gene nucleotide mutation G22017T, the prevalence of which increased from $\leq 5\%$ in the first sample to $\geq 60\%$ in the second sample. This mutation corresponds to amino acid substitution W152L within the N-terminal domain of the spike protein, which interacts with neutralizing antibodies. Siquiera et al. studied intrahost genetic diversity in nasopharyngeal and oropharyngeal samples from 57 SARS-CoV-2-positive cancer patients and 14 SARS-CoV-2-positive healthcare workers from the Brazilian National Cancer Institute who were sampled between April and May 2020.¹¹¹ In addition to the majority of mutations, 85 other mutations were detected at frequencies varying between 1.4% and 19.7%. The cancer patients exhibited significantly higher intrahost viral genetic diversity than the healthcare workers. Dudouet et al. performed next-generation sequencing with Nanopore technology of SARS-CoV-2 genomes from 78 patients with nasopharyngeal virus persistence beyond 17 days and from 47 patients without such virus persistence.¹¹² They reported that intrahost diversity was higher in the persistent than nonpersistent patients, with minority N501Y and P681H mutations in the patients of both groups. Caccuri et al. described the case of a 59-year-old male patient who was immunocompromised due to follicular lymphoma and chronically infected with SARS-CoV-2 for 222 days.¹¹³ They reported progressive emergence of the majority quasispecies of viruses with critical spike mutations Q493K and N501T, but that the prevalence of quasispecies with deletion of spike amino acid 144 fluctuated over time. In another work, Sonleitner

et al. studied the mutational dynamics of SARS-CoV-2 in serial passages *in vitro*.¹¹⁴ They used 10 ancestral SARS-CoV-2 strains for consecutive serial incubation in 10 passages in Vero B4 cells, observing that the spike gene was a hot spot for mutations, including substitutions at codons 484 and 501, which are backbone mutations of several variants of concern (<https://covariants.org/>).¹¹⁵

The therapeutic use of monoclonal antibodies was reported to have a significant impact on the composition of SARS-CoV-2 quasispecies by promoting the emergence of quasispecies that are less susceptible to this therapy. Peiffer-Smadja et al. and Jensen et al. detected the emergence of spike E484K-harboring SARS-CoV-2 in 10 of 12 patients treated with bamlanivimab using next-generation sequencing.^{116,117} Lohr et al. reported a case of emergence of spike E484K-harboring SARS-CoV-2 of the Alpha variant after bamlanivimab therapy in a patient with acute myeloid leukemia.¹¹⁸ Additionally, Guigon et al. detected the spike mutation Q493R in a 63-year-old patient with cutaneous T-cell lymphoma at Day 40 of bamlanivimab/etesevimab administration that was deemed compatible with reduced efficacy or lack of efficacy of the monoclonal antibodies.¹¹⁹ A study by Vellas et al. involved the use of the PacBio SMRT system to examine the intrahost diversity of the spike gene in 32 patients who received bamlanivimab alone (4 patients), bamlanivimab/etesevimab (23 patients), or cambirivimab/imdevimab (5 patients).¹²⁰ These patients included 11 solid organ transplant recipients and 6 patients who were immunocompromised due to other causes. SARS-CoV-2 with the spike protein Q493R mutation was detected in two cases on Day 7 post-administration of bamlanivimab/etesevimab and in one case on Day 14. In a fourth case, the Q493K spike mutation was detected on Day 7, and in a fifth case, the E484K spike mutation was detected on Day 21. These findings were associated in three cases with the absence of viral load decrease or with viral rebound. In contrast, in another work, in 36 patients who were administered casirivimab/imdevimab, no spike mutations were detected in the SARS-CoV-2 quasispecies associated with reduced monoclonal antibody activity.¹²¹

In addition, compartmentalization of SARS-CoV-2 quasispecies has been reported in a few studies. Jary et al. performed next-generation sequencing on PCR amplification products for genes encoding the spike, envelope, membrane, and nucleocapsid proteins in a SARS-CoV-2-infected patient, and minority viral populations represented up to 1% of the whole population during the course of infection.¹²² Quasispecies differed in sequential samples and between lower and upper respiratory tract samples collected on the same day, suggesting independent sites of SARS-CoV-2 replication. By studying samples of extrapulmonary organs, including the heart, kidney, liver, and spleen, Van Cleemput et al. described organ-specific intrasample diversity in 13 postmortem cases involving immunocompromised patients.⁸⁵ Strikingly, they identified the spike mutation of concern N501Y, which mostly emerged in the Alpha, Beta, and Gamma variants worldwide during late 2020 and early 2021, though the 13 patients had been sampled in March 2020. In addition, the spike substitution T1027I present in the Gamma variant and Y453F, initially described in mink,¹²³ were detected as minority

quasispecies in several organs. Hartard et al. performed a post-mortem study of multiple specimens collected from a patient who revealed heterogeneous distribution of viral quasispecies according to the tissue examined, including V341A, E654A, and H655R substitutions in spike.¹²⁴ Additionally, Gaiarsa et al. described quasispecies based on the S gene in 109 clinical samples collected from 77 patients, including from the upper respiratory tract and lower respiratory tract.¹²⁵ They identified a greater incidence among minority quasispecies of nonsynonymous mutations and indels in the lower respiratory tract, which they asserted could be the result of an ability of the virus to invade cells without interacting with ACE2. In addition, they observed that intrasample diversity was more prevalent in the gene region encoding the spike cleavage site, particularly in samples from the upper respiratory tract. Also, an association of five mutations with the upper respiratory tract (U2055G, U2058G, U2060G, U2100G, and A3483S) and four mutations with the lower respiratory tract (C212A, A3005U, C3485U/G, and A3596G/U) was found. Finally, Normandin et al. performed next-generation sequencing on 120 specimens using formalin-fixed, paraffin-embedded tissues from six patients who experienced fatal COVID-19.¹²⁶ They reported that a majority of the 180 SARS-CoV-2 mutational patterns identified in these patients were unique to given tissue samples, indicating compartmentalized infections in extrapulmonary sites. The spike protein amino acid substitution Q675H, present as a minority mutation in heart tissue, was not circulating at the time the patient died but was later detected in the consensus genomes from several SARS-CoV-2 lineages. In addition, a prevalence of minority patterns greater than 10% was more frequently detected in patients who experienced a long duration of COVID-19 before death.

5.3 | Interactions between SARS-CoV-2 and host cells and the quasispecies paradigm

The plasma membrane of host cells acts as a machine for concentrating viruses. Two mechanisms are at work: (i) a reduction in dimensionality due to the transition from a three-dimensional space containing the viruses to the two-dimensional space of the membrane, which favors the concentration of the viruses on the surface of target cells¹²⁷; and (ii) the presence on the plasma membrane of lipid rafts (microdomains enriched in gangliosides and cholesterol), which are the preferred sites for the landing of most viruses on host cells.¹²⁸ In fact, regardless of the target tissue or cell, rafts display a negative electrostatic field that attracts viral particles, the surfaces of which are studded with electropositive zones.¹²⁹ Viruses from the “wild bunch” of an inoculum, which have already undergone selection in the previous host, necessarily fulfill these conditions and will therefore be globally attracted by target cells. It is only after this initial interaction with lipid rafts that more specific interactions involving protein receptors expressed by the target cells can occur. It is nevertheless striking to note that these latter interactions are based on recognition by the virus not of a single

receptor but of several. For instance, in the case of SARS-CoV-2, several membrane receptors have been identified: ACE-2 (in the respiratory tract, lung, brain, gastrointestinal tract, blood vessels, heart, kidneys, testis); β 1 and β 3 integrins (on pulmonary epithelial cells); GRP78 (thyroid gland, olfactory cells, lung macrophages); CD26 (CD4+ T cells, myeloid cells, myocardial cells, blood capillaries); AXL (pulmonary epithelium); CD147 (neural cells, lymph nodes, platelets, red blood cells, myeloid cells, skin); NRP-1 (olfactory epithelium, bone marrow-derived macrophages); and CD209L/CD209 lectins (human endothelium, renal tissue, megakaryocytes, platelets).¹²⁹ This extreme diversity of protein receptors and coreceptors provides a broad range of possibilities for the “wild bunch,” ensuring rapid dissemination of viruses in the host and increasing virus genetic diversity as a consequence of adaptation to the receptor.¹³⁰ From a molecular point of view, it is also of interest to note that the region of the SARS-CoV-2 spike protein that faces the host cell membrane is organized into two separate domains: the N-terminal domain (NTD), which mediates attachment of the virus to negatively charged heparan sulfates and lipid raft gangliosides; and the receptor binding domain, which is initially masked at the center of the trimeric viral spike and becomes accessible to ACE-2 and alternative receptors after conformational reorganization of the spike trimer.¹³¹ These geometrical features clearly optimize dissemination of the virus throughout the host body by allowing attachment of the virus to a wide range of cells and then the likelihood of finding a compatible receptor on different cell types. Overall, the molecular biology of host–virus interactions is therefore essentially compatible with the “wild bunch” concept with the attraction of viruses in an inoculum to lipid rafts according to the double facilitation mechanism (reduction of dimensionality/electrostatic attraction) and the biochemical diversity of viral protein receptors. In the case of SARS-CoV-2, the wide diversity of symptoms in various organs occurring during the acute phase of the infection is also consistent with the “wild bunch” concept.

6 | QUASISPECIES SELECTION

In RNA viruses in particular, the frequency and importance of quasispecies leads, as described by Domingo and Perales, to adaptive pluripotency,⁴⁸ without discussing the probably quite important role in the evolution of recombinant RNA viruses and evidenced by enteroviruses and SARS-CoV-2.^{3,4} The large number of quasispecies is probably at the origin of the rebounds in virulence or transmissibility that have been postulated not to occur according to the Muller–Ratchet principle.¹³² In this theory, the lack of recombination can lead to a gradual decrease in fitness and extinction. It is likely that this phenomenon is at the origin of the usual epidemic curves of acutely infecting RNA viruses, which exhibit bell-shaped curves, appearing after a favorable mutation and gradually decreasing due to accumulation of neutral or unfavorable mutations.¹³³ We have evaluated for SARS-CoV-2 the number of acceptable unfavorable mutations, which is on the order of 6–7, leading to extinction of the

“democratic” genome clone if no new favorable mutations appear.¹³³ These favorable mutations exist in quasispecies long before they are found in epidemic “democratic” genomes, indicating the existence of a large phenotypic reservoir in RNA viruses. The benefits of conserving this large diversity can indeed lead to selection according to the circumstances. In the case of SARS-CoV-2, the rapid use of monoclonal antibodies has favored the emergence of minority components of quasispecies that have become “democratic” genomes under the effect of their selection pressure. This phenomenon has been observed in new viruses with various antiviral treatments.¹³⁴ Vaccines also have the ability to select mutations that change the majority target of the immune response.⁴³ All these elements highlight that for antiviral therapeutic strategies, the choice of a narrow target entails the risk of emergence of an unknown variant existing only as a minority population in quasispecies.

A less commonly discussed fact is the opposite effect, whereby lower virulence allows for longer persistence. This has been referred to as the “flattest advantage.”^{48,135} It appears that less aggressive mutants are less likely to decline rapidly in activity due to counterselection, which may be related to the host response or to the stronger effect of favoring sequences in the more active population. Finally, as in some viral infections more than 10^{12} viruses can be quantified, the quasispecies present at any one time of infection is absolutely paramount and lead to an understanding of why it is so difficult to predict evolution of epidemics of novel viruses, especially RNA viruses.

7 | FINDING BLACKLISTED MUTATIONS

Overall, the analysis of the mutations produced in circulating SARS-CoV-2 genomes shows a very large variety of mutations. In a recent work, we highlighted 22 225 nucleotide mutations observed in a library of 61 397 genomes produced in our institute and available in databases.¹³³ These observed mutations cannot, by definition, be lethal because lethal mutations are not followed by reproduction. We have separated these types of mutations into “hyperfertile,” weakly “fertile,” neutral or weakly deleterious, or absent. Other authors have classified mutations as “whitelist” mutations or “blacklist” mutations, which are never found and assumed to be lethal.¹³⁶ The analysis of lethal mutations is interesting; at the moment, such analysis can probably only be conducted in the real world through the study of quasispecies. Quasispecies analysis will reveal whether the distribution of lethal mutations is random (in the Court Jester model^{133,137}) or whether certain lethal mutations are more frequently represented than would be expected by chance, as has been observed for beneficial mutations. Some beneficial mutations are found as much more represented, especially in spike; conversely, some regions have very few beneficial mutations, especially regions containing genes essential for replication, such as that encoding the RNA polymerase.¹³³ It will be interesting to evaluate the frequency of point mutations never expressed in transmitted genomes to identify particular areas where such mutations originate as a probable

consequence of the host. It is clear, for example, that changes related to apolipoprotein-B (ApoB) mRNA editing enzyme catalytic polypeptide-like protein (APOBEC) activity¹³⁸ should be more common in this type of mutation, as APOBEC function is effectively to neutralize viral activity by replacing one base with another through deamination. As a matter of fact, some studies have shown that the sequenced viruses containing the most nonreproducible quasispecies are more likely to be associated with mild than severe forms of the disease.^{139–141} It will be interesting to assess the nature of the changes observed and their frequency among unexpressed mutations to determine if other host-related mechanisms of virus neutralization exist. This might reflect an incomplete resistance mechanism that reduces the severity of the disease. Thus, specific study of the frequency of lethal unexpressed mutations and their percentage may be an indicator of the efficiency of an immune defense that is little or poorly explored in the context of acute viral infections.

8 | CONCLUSION

Current knowledge is evolving extremely rapidly due to the explosion of the number of available viral genomes. There was once definitely (apart from the major virological fields such as HIV or viral hepatitis) a delay in the number of sequences available. For example, if we compare the number of sequences available for *Escherichia coli*, *Klebsiella pneumoniae*, or *Salmonella typhi* and the number of sequences available for influenza, measles, or chickenpox viruses, we can only be astonished by the lack of data regarding viruses. Currently, available data, particularly in the field of measles,⁸² show that we must definitely move away from the concept of the virus as a single object, as a clone represented by a single sequence. The habit of representing sequences by a single molecule is hardly acceptable, given that with current tools, we know that in most cases there are quasispecies in both natural samples and viral cultures. Interestingly, work on measles viruses has shown that not only these quasispecies develop during infection but that infection itself can occur due to quasispecies with different organ affinities. Thus, partly as a result of the SARS-CoV-2 pandemic for which more than 15 million “democratic” genome sequences are now available and for which studies and attempts to understand the different epidemic waves caused by distinct variants as well as their host and even organ affinity have been paramount, we should turn a page in the understanding of viral infections and viruses. Viruses are not inert biomolecules with a rare mutation but organisms with many variations, with different mutants or variants causing infections.

In fact, we realize that almost all published virus genome sequences are approximate and do not represent a biological reality. It is not possible to define a virus by a single sequence as the “democratic” sequence, and the actual experience of deep sequencing reveals huge diversity when several thousand sequences are generated from a given genomic region. This suggests that the genomes currently available are merely a representation of reality underlining the most common nucleotide at each site of the

genomes but not representing an authentic reality. Figures 1C and 2 indicate that in the patient’s blood or respiratory sample, many sequence reads generated by next-generation sequencing, despite being very short (approximately 50 nucleotides long), differ from the “democratic” sequence because they have mutations relative to the virtual sequence. Clearly, this means that even the notion of quasispecies is not precise enough. The existence of a mutation relative to a virtual sequence is compelling. However, only accumulation of mutations at the same genomic sites in several sequence reads can lead to the definition of quasispecies, which does not prevent the appearance of mutations at sites other than those that define the quasispecies. In practice, the amazing amounts of sequence data accumulated for HIV and more recently for SARS-CoV-2 are leading us to radically rethink viruses as a wild bunch of virus particles shuttling related but different genomic sequences; they are constantly innovating and producing epidemics and pathological episodes of a complexity that is difficult to analyze. This also calls into question in a very significant way the possibilities of creating a single genome representing a specific invariant pathotype or creating RNA vaccines with a unique sequence, as was shown in various studies.¹⁵¹ It is therefore becoming essential for databanks (as has started to occur) to not only make available the majority, or “democratic” genomes but also all quasispecies as supplementary data so that they can be analyzed. The technical tools needed to analyze such databases, including quasispecies, should be developed in the coming years to decipher these data in this completely new era of virology.

AUTHOR CONTRIBUTIONS

Didier Raoult and Philippe Colson designed the study. All authors wrote, reviewed, and approved the final manuscript.

ACKNOWLEDGMENTS

This study was supported by the French Government under the “Investments for the Future” program managed by the National Agency for Research (ANR) (Méditerranée-Infection 10-IAHU-03). We thank Jérémy Delerce for his help with the conception of the figures.

CONFLICT OF INTEREST STATEMENT

Didier Raoult declares grants or contracts and royalties or licenses from Hitachi High-Technologies Corporation, Tokyo, Japan. He is a scientific board member of Eurofins company and a founder and shareholder of a microbial culture company (Culture Top), two biotechnology companies (Techno-Jouvence, and Gene and Green TK), and an infectious diseases rapid diagnosis company (Pocramé). The remaining authors declare no conflict of interest. Funding sources played no role in the design and conduct of the study, and in the preparation, review, or approval of the manuscript.

DATA AVAILABILITY STATEMENT

Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

ORCID

Philippe Colson  <http://orcid.org/0000-0001-6285-0308>

Didier Raoult  <http://orcid.org/0000-0002-0633-5974>

REFERENCES

- Raoult D, Forterre P. Redefining viruses: lessons from Mimivirus. *Nat Rev Microbiol*. 2008;6:315-319. doi:10.1038/nrmicro1858
- Pérez-Losada M, Arenas M, Galán JC, Palero F, González-Candelas F. Recombination in viruses: mechanisms, methods of study, and evolutionary consequences. *Infect Genet Evol*. 2015;30:296-307. doi:10.1016/j.meegid.2014.12.022
- Amona I, Medkour H, Akiana J, et al. Enteroviruses from humans and great apes in the Republic of Congo: recombination within enterovirus C serotypes. *Microorganisms*. 2020;8:1779. doi:10.3390/microorganisms8111779
- Colson P, Fournier PE, Delerce J, et al. Culture and identification of a "Deltamicro" SARS-CoV-2 in a three cases cluster in southern France. *J Med Virol*. 2022;94:3739-3749. doi:10.1002/jmv.27789
- Baptiste E, Lopez P, Bouchard F, Baquero F, McInerney JO, Burian RM. Evolutionary analyses of non-genealogical bonds produced by introgressive descent. *Proc Natl Acad Sci USA*. 2012;109:18266-18272. doi:10.1073/pnas.1206541109
- Dagan T, Martin W. The tree of one percent. *Genome Biol*. 2006;7:118. doi:10.1186/gb-2006-7-10-118
- Raoult D. The post-darwinist rhizome of life. *Lancet*. 2010;375:104-105. doi:10.1016/S0140-6736(09)61958-9
- Ibrahim A, Colson P, Merhej V, et al. Rhizomal reclassification of living organisms. *Int J Mol Sci*. 2021;22:5643. doi:10.3390/ijms22115643
- Ivanovski D. Concerning the mosaic disease of tobacco plant. *St Petersburg Acad Imp Sci Bul*. 1892;35:67-70.
- Beijerinck MW. Concerning a contagium vivum fluidum as a cause of the spot-disease of tobacco leaves. *Verh Akad Wet Amster*. 1898;6:33-55.
- Bawden FC, Pirie NW, Bernal JD, Fankuchen I. Liquid crystalline substances from virus infected plants. *Nature*. 1936;138:1051-1052.
- Fraenkel-Conrat H, Singer B, Williams RC. Infectivity of viral nucleic acid. *Biochim Biophys Acta*. 1957;25:87-96. doi:10.1016/0006-3002(57)90422-5
- Stanley WM. Isolation of a crystalline protein possessing the properties of tobacco mosaic virus. *Science*. 1935;81:644-645.
- Burnet MF. *The Virus and the Cell*. The second R. E. Dyer Lecture at the National Institutes of Health, October 29, 1952. U.S. Govt. Printing Office [U.S. Department of Health, Education and Welfare, PHS Publ. no. 328, NIH]; 1952: 19.
- Lwoff A. The concept of virus. *J Gen Microbiol*. 1957;17:239-253. doi:10.1099/00221287-17-2-239
- Fauquet CM, Stanley J. Revising the way we conceive and name viruses below the species level: a review of geminivirus taxonomy calls for new standardized isolate descriptors. *Arch Virol*. 2005;150:2151-2179. doi:10.1007/s00705-005-0583-0
- Suárez NM, Wilkie GS, Hage E, et al. Human cytomegalovirus genomes sequenced directly from clinical material: variation, multiple-strain infection, recombination, and gene loss. *J Infect Dis*. 2019;220:781-791. doi:10.1093/infdis/jiz208
- Iyer LM, Balaji S, Koonin EV, Aravind L. Evolutionary genomics of nucleo-cytoplasmic large DNA viruses. *Virus Res*. 2006;117:156-184. doi:10.1016/j.virusres.2006.01.009
- Senkevich TG, Yutin N, Wolf YI, Koonin EV, Moss B. Ancient gene capture and recent gene loss shape the evolution of orthopoxvirus-host interaction genes. *mBio*. 2021;12:e0149521. doi:10.1128/mBio.01495-21
- Boyer M, Azza S, Barrassi L, et al. Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc Natl Acad Sci USA*. 2011;108:10296-10301. doi:10.1073/pnas.1101181108
- Colson P, Raoult D. Lamarckian evolution of the giant Mimivirus in allopatric laboratory culture on amoebae. *Front Cell Infect Microbiol*. 2012;2:91. doi:10.3389/fcimb.2012.00091
- Pasteur L. De l'atténuation du virus du Choléra des poules. *C R Hebd Seanc Acad Sci*. 1880;91:673-680.
- Darwin C., Keblet L. *On the Origin of Species by Means of Natural Selection, or, the Preservation of Favoured Races in the Struggle for Life*. J. Murray; 1859.
- Fiers W, Contreras R, Duerinck F, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*. 1976;260:500-507. doi:10.1038/260500a0
- Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA*. 1977;74:5463-5467. doi:10.1073/pnas.74.12.5463
- Eigen M. On the nature of virus quasispecies. *Trends Microbiol*. 1996;4:216-218. doi:10.1016/0966-842X(96)20011-3
- Siddell SG, Smith DB, Adriaenssens E, et al. Virus taxonomy and the role of the International Committee on Taxonomy of Viruses (ICTV). *J Gen Virol*. 2023;104:001840. doi:10.1099/jgv.0.001840
- International Committee on Taxonomy of Viruses Executive Committee. The new scope of virus taxonomy: partitioning the virosphere into 15 hierarchical ranks. *Nat Microbiol*. 2020;5:668-674. doi:10.1038/s41564-020-0709-x
- Simmonds P, Adriaenssens EM, Zerbini FM, et al. Four principles to establish a universal virus taxonomy. *PLoS Biol*. 2023;21:e3001922. doi:10.1371/journal.pbio.3001922
- Findlay GM. Variation in animal viruses. A review. *J R Microsc Soc*. 1936;56:213-299.
- Burnet F, Bull DH. Changes in influenza virus associated with adaptation to passage in chick embryos. *Aust J Exp Biol Med Sci*. 1943;21:55-69.
- Burnet MF. *Virus as Organism. Evolutionary And Ecological Aspects of Some Human Viral Diseases [Dunham Lectures, Harvard University, 1944]*. Harvard University Press;1945: 9.
- Granoff A. Induction of newcastle disease virus mutants with nitrous acid. *Virology*. 1961;13:402-408. doi:10.1016/0042-6822(61)90270-7
- Eggers HJ, Tamm I. Coxsackie A9 virus: mutation from drug dependence to drug independence. *Science*. 1965;148:97-98. doi:10.1126/science.148.3666.97
- Pringle CR. Genetic characteristics of conditional lethal mutants of vesicular stomatitis virus induced by 5-fluorouracil, 5-azacytidine, and ethyl methane sulfonate. *J Virol*. 1970;5:559-567. doi:10.1128/JVI.5.5.559-567.1970
- Holland JJ, Grabau EA, Jones CL, Semler BL. Evolution of multiple genome mutations during long-term persistent infection by vesicular stomatitis virus. *Cell*. 1979;16:495-504. doi:10.1016/0092-8674(79)90024-2
- Batschelet E, Domingo E, Weissmann C. The proportion of revertant and mutant phage in a growing population, as a function of mutation and growth rate. *Gene*. 1976;1:27-32. doi:10.1016/0378-1119(76)90004-4
- Domingo E, Sabo D, Taniguchi T, Weissmann C. Nucleotide sequence heterogeneity of an RNA phage population. *Cell*. 1978;13:735-744. doi:10.1016/0092-8674(78)90223-4
- Domingo E, Dávila M, Ortín J. Nucleotide sequence heterogeneity of the RNA from a natural population of foot-and-mouth-disease virus. *Gene*. 1980;11:333-346. doi:10.1016/0378-1119(80)90073-6
- Sanger F, Coulson AR, Hong GF, Hill DF, Petersen GB. Nucleotide sequence of bacteriophage λ DNA. *J Mol Biol*. 1982;162:729-773. doi:10.1016/0022-2836(82)90546-0

41. Rothberg JM, Leamon JH. The development and impact of 454 sequencing. *Nat Biotechnol.* 2008;26:1117-1124. doi:10.1038/nbt1485
42. Mokili JL, Rohwer F, Dutilh BE. Metagenomics and future perspectives in virus discovery. *Curr Opin Virol.* 2012;2:63-77. doi:10.1016/j.coviro.2011.12.004
43. Carabelli AM, Peacock TP, Thorne LG, et al. SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nat Rev Microbiol.* 2023;21:162-177. doi:10.1038/s41579-022-00841-7
44. Eigen M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften.* 1971;58:465-523. doi:10.1007/BF00623322
45. Eigen M, Schuster P. A principle of natural self-organization: part A: emergence of the hypercycle. *Naturwissenschaften.* 1977;64:541-565. doi:10.1007/BF00450633
46. Biebricher CK, Eigen M. What is a quasispecies? *Current Topics in Microbiology and Immunology.* Vol 299, 2006 ed., 2006:1-31. Springer Verlag.
47. Domingo E. Quasispecies and the implications for virus persistence and escape. *Clin Diagn Virol.* 1998;10:97-101. doi:10.1016/s0928-0197(98)00032-4
48. Domingo E, Perales C. Viral quasispecies. *PLoS Genet.* 2019;15:e1008271. doi:10.1371/journal.pgen.1008271
49. Jones LR, Zandomeni R, Weber EL. Quasispecies in the 5' untranslated genomic region of bovine viral diarrhoea virus from a single individual. *J Gen Virol.* 2002;83:2161-2168. doi:10.1099/0022-1317-83-9-2161
50. Jerzak G, Bernard KA, Kramer LD, Ebel GD. Genetic variation in west Nile virus from naturally infected mosquitoes and birds suggests quasispecies structure and strong purifying selection. *J Gen Virol.* 2005;86:2175-2183. doi:10.1099/vir.0.81015-0
51. Herring BL, Tsui R, Peddada L, Busch M, Delwart EL. Wide range of quasispecies diversity during primary hepatitis C virus infection. *J Virol.* 2005;79:4340-4346. doi:10.1128/JVI.79.7.4340-4346.2005
52. Costa-Mattioli M, Domingo E, Cristina J. Analysis of sequential hepatitis A virus strains reveals coexistence of distinct viral subpopulations. *J Gen Virol.* 2006;87:115-118. doi:10.1099/vir.0.81286-0
53. Kassaye S, Lee E, Kantor R, et al. Drug resistance in plasma and breast milk after single-dose nevirapine in subtype C HIV type 1: population and clonal sequence analysis. *AIDS Res Hum Retroviruses.* 2007;23:1055-1061. doi:10.1089/aid.2007.0045
54. Sun F, Wang X, Tan S, et al. SARS-CoV-2 quasispecies provides an advantage mutation pool for the epidemic variants. *Microbiol Spectr.* 2021;9:e0026121. doi:10.1128/Spectrum.00261-21
55. Domingo E, Sheldon J, Perales C. Viral quasispecies evolution. *Microbiol Mol Biol Rev.* 2012;76:159-216. doi:10.1128/MMBR.05023-11
56. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature.* 2006;439:344-348. doi:10.1038/nature04388
57. Lipsitch M, O'Hagan JJ. Patterns of antigenic diversity and the mechanisms that maintain them. *J R Soc Interface.* 2007;4:787-802. doi:10.1098/rsif.2007.0229
58. Echevarria JM, Avellón A. Improved detection of natural hepatitis B virus surface antigen (HbsAg) mutants by a new version of the VITROS HbsAg assay. *J Med Virol.* 2008;80:598-602. doi:10.1002/jmv.21146
59. Pfeiffer JK, Kirkegaard K. Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. *Proc Natl Acad Sci USA.* 2006;103:5520-5525. doi:10.1073/pnas.0600834103
60. Borrego B, Novella IS, Giralto E, Andreu D, Domingo E. Distinct repertoire of antigenic variants of foot-and-mouth disease virus in the presence or absence of immune selection. *J Virol.* 1993;67:6071-6079. doi:10.1128/JVI.67.10.6071-6079.1993
61. Chumakov KM, Powers LB, Noonan KE, Roninson IB, Levenbook IS. Correlation between amount of virus with altered nucleotide sequence and the monkey test for acceptability of oral poliovirus vaccine. *Proc Natl Acad Sci USA.* 1991;88:199-203. doi:10.1073/pnas.88.1.199
62. Holland JJ, De La Torre JC, Steinhauer DA. RNA virus populations as quasispecies. *Curr Top Microbiol Immunol.* 1992;176:1-20. doi:10.1007/978-3-642-77011-1_1
63. Eddy SR. Profile hidden Markov models. *Bioinformatics.* 1998;14:755-763. doi:10.1093/bioinformatics/14.9.755
64. Hirsch MS, Günthard HF, Schapiro JM, et al. Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Clin Infect Dis.* 2008;47:266-285. doi:10.1086/589297
65. Mellors JW, Rinaldo Jr., CR, Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science.* 1996;272:1167-1170. doi:10.1126/science.272.5265.1167
66. Larder BA, Kemp SD, Purifoy DJ. Infectious potential of human immunodeficiency virus type 1 reverse transcriptase mutants with altered inhibitor sensitivity. *Proc Natl Acad Sci USA.* 1989;86:4803-4807. doi:10.1073/pnas.86.13.4803
67. St Clair MH, Martin JL, Tudor-Williams G, et al. Resistance to ddI and sensitivity to AZT induced by a mutation in HIV-1 reverse transcriptase. *Science.* 1991;253:1557-1559. doi:10.1126/science.1716788
68. Katzenstein DA. Viral phenotype and genotype as markers in clinical trials. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1995;10(suppl 2):S25-S34.
69. Hirsch MS, Conway B, D'Aquila RT, et al. Antiretroviral drug resistance testing in adults with HIV infection: implications for clinical management. *JAMA.* 1998;279:1984-1991. doi:10.1001/jama.279.24.1984
70. Ho CKY, Welkers MRA, Thomas XV, et al. A comparison of 454 sequencing and clonal sequencing for the characterization of hepatitis C virus NS3 variants. *J Virol Methods.* 2015;219:28-37. doi:10.1016/j.jviromet.2015.03.018
71. Boyer JC, Bebenek K, Kunkel TA. Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc Natl Acad Sci USA.* 1992;89:6919-6923. doi:10.1073/pnas.89.15.6919
72. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol.* 2010 84:9864-9878. doi:10.1128/JVI.00915-10
73. Wang YM, Dyer WB, Workman C, Wang B, Sullivan JS, Saksena NK. Molecular evidence for drug-induced compartmentalization of HIV-1 quasispecies in a patient with periodic changes to HAART. *AIDS.* 2000;14:2265-2272. doi:10.1097/00002030-200010200-00007
74. Pinggen M, Nijhuis M, de Bruijn JA, Boucher CAB, Wensing AMJ. Evolutionary pathways of transmitted drug-resistant HIV-1. *J Antimicrob Chemother.* 2011;66:1467-1480. doi:10.1093/jac/dkr157
75. Metzner KJ, Scherrer AU, Preiswerk B, et al. Origin of minority drug-resistant HIV-1 variants in primary HIV-1 infection. *J Infect Dis.* 2013;208:1102-1112. doi:10.1093/infdis/jit310
76. Tamalet C, Yahi N, Tourrés C, et al. Multidrug resistance genotypes (insertions in the β 3- β 4 Finger subdomain and MDR mutations) of HIV-1 reverse transcriptase from extensively treated patients:

- incidence and association with other resistance mutations. *Virology*. 2000;270:310-316. doi:10.1006/viro.2000.0261
77. Paredes R, Lalama CM, Ribaudo HJ, et al. Pre-existing minority drug-resistant HIV-1 variants, adherence, and risk of antiretroviral treatment failure. *J Infect Dis*. 2010;201:662-671. doi:10.1086/650543
 78. Lederberg J, Lederberg EM. Replica plating and indirect selection of bacterial mutantS. *J Bacteriol*. 1952;63:399-406. doi:10.1128/jb.63.3.399-406.1952
 79. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform*. 2013;14:178-192. doi:10.1093/bib/bbs017
 80. Bellecave P, Recordon-Pinson P, Papuchon J, et al. Detection of low-frequency HIV type 1 reverse transcriptase drug resistance mutations by ultradeep sequencing in naive HIV type 1-infected individuals. *AIDS Res Hum Retroviruses*. 2014;30:170-173. doi:10.1089/AID.2013.0130
 81. Balakrishna S, Loosli T, Zaheri M, et al. Frequency matters: comparison of drug resistance mutation detection by Sanger and next-generation sequencing in HIV-1. *J Antimicrob Chemother*. 2023;78:656-664. doi:10.1093/jac/dkac430
 82. Cattaneo R, Donohue RC, Generous AR, Navaratnarajah CK, Pfaller CK. Stronger together: multi-genome transmission of measles virus. *Virus Res*. 2019;265:74-79. doi:10.1016/j.virusres.2019.03.007
 83. Donohue RC, Pfaller CK, Cattaneo R. Cyclical adaptation of measles virus quasispecies to epithelial and lymphocytic cells: to V, or not to V. *PLoS Pathog*. 2019;15:e1007605. doi:10.1371/journal.ppat.1007605
 84. Xiao Y, Dolan PT, Goldstein EF, et al. Poliovirus intrahost evolution is required to overcome tissue-specific innate immune responses. *Nat Commun*. 2017;8:375. doi:10.1038/s41467-017-00354-5
 85. Van Cleemput J, van Snippenberg W, Lambrechts L, et al. Organ-specific genome diversity of replication-competent SARS-CoV-2. *Nat Commun*. 2021;12:6612. doi:10.1038/s41467-021-26884-7
 86. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579:270-273. doi:10.1038/s41586-020-2012-7
 87. Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. *Acta Biomed*. 2020;91:157-160. doi:10.23750/abm.v91i1.9397
 88. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet*. 2020;395:565-574. doi:10.1016/S0140-6736(20)30251-8
 89. Koyama T, Platt D, Parida L. Variant analysis of SARS-CoV-2 genomes. *Bull World Health Organ*. 2020;98:495-504. doi:10.2471/BLT.20.253591
 90. Wang S, Xu X, Wei C, et al. Molecular evolutionary characteristics of SARS-CoV-2 emerging in the United States. *J Med Virol*. 2022;94:310-317. doi:10.1002/jmv.27331
 91. Van Dorp L, Richard D, Tan CCS, Shaw LP, Acman M, Ballouf F. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. *Nat Commun*. 2020;11:5986. doi:10.1038/s41467-020-19818-2
 92. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018;34:4121-4123. doi:10.1093/bioinformatics/bty407
 93. Kim K, Calabrese P, Wang S, et al. The roles of APOBEC-mediated RNA editing in SARS-CoV-2 mutations, replication and fitness. *Sci Rep*. 2022;12:14972. doi:10.1038/s41598-022-19067-x
 94. Ringlander J, Fingal J, Kann H, et al. Impact of ADAR-induced editing of minor viral RNA populations on replication and transmission of SARS-CoV-2. *Proc Natl Acad Sci USA*. 2022;119:e2112663119. doi:10.1073/pnas.2112663119
 95. Gribble J, Stevens LJ, Agostini ML, et al. The coronavirus proof-reading exoribonuclease mediates extensive viral recombination. *PLoS Pathog*. 2021;17:e1009226. doi:10.1371/journal.ppat.1009226
 96. Jackson B, Boni MF, Bull MJ, et al. Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell*. 2021;184:5179-5188.e8. doi:10.1016/j.cell.2021.08.014
 97. Adami C, Pooley J, Glomb J, et al. Evolution of mouse hepatitis virus (MHV) during chronic infection: quasispecies nature of the persisting MHV RNA. *Virology*. 1995;209:337-346. doi:10.1006/viro.1995.1265
 98. Rowe CL, Fleming JO, Nathan MJ, Sgro JY, Palmenberg AC, Baker SC. Generation of coronavirus spike deletion variants by high-frequency recombination at regions of predicted RNA secondary structure. *J Virol*. 1997;71:6183-6190. doi:10.1128/JVI.71.8.6183-6190.1997
 99. Rowe CL, Baker SC, Nathan MJ, Sgro JY, Palmenberg AC, Fleming JO. Quasispecies development by high frequency RNA recombination during MHV persistence. In: Enjuanes L, Siddell SG, Spaan W, eds. *Coronaviruses and Arteriviruses* (Advances in Experimental Medicine and Biology; Vol. 440). Springer US; 1998:759-765. http://link.springer.com/10.1007/978-1-4615-5331-1_98
 100. Kiss I, Kecskeméti S, Tanyi J, Belák S, Ros C, B. Klingeborn S. Observations on the quasispecies composition of three animal pathogenic RNA viruses. *Acta Vet Hung*. 1999;47:471-480. doi:10.1556/Avet.47.1999.4.7
 101. Zhang X, Hasoksuz M, Spiro D, et al. Quasispecies of bovine enteric and respiratory coronaviruses based on complete genome sequences and genetic changes after tissue culture adaptation. *Virology*. 2007;363:1-10. doi:10.1016/j.virol.2007.03.018
 102. Woo PC, Lau SK, Tsang CC, et al. Coronavirus HKU15 in respiratory tract of pigs and first discovery of coronavirus quasispecies in 5'-untranslated region. *Emerg Microbes Infect*. 2017;6:e53. doi:10.1038/emi.2017.37
 103. Viehweger A, Krautwurst S, Lamkiewicz K, et al. Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res*. 2019;29:1545-1554. doi:10.1101/gr.247064.118
 104. Gorse GJ, Patel GB, Fan X. Interpatient mutational spectrum of human coronavirus-OC43 revealed by illumina sequencing. *J Med Virol*. 2017;89:1330-1338. doi:10.1002/jmv.24780
 105. Lu X, Rowe LA, Frace M, et al. Spike gene deletion quasispecies in serum of patient with acute MERS-CoV infection. *J Med Virol*. 2017;89:542-545. doi:10.1002/jmv.24652
 106. Sender R, Bar-On YM, Gleizer S, et al. The total number and mass of SARS-CoV-2 virions. *Proc Natl Acad Sci USA*. 2021;118:e2024815118. doi:10.1073/pnas.2024815118
 107. Choi B, Choudhary MC, Regan J, et al. Persistence and evolution of SARS-CoV-2 in an immunocompromised host. *N Engl J Med*. 2020;383:2291-2293. doi:10.1056/NEJMc2031364
 108. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. *Nature*. 2021;592:277-282. doi:10.1038/s41586-021-03291-y
 109. Laubscher F, Cordey S, Friedlaender A, et al. SARS-CoV-2 evolution among oncological population: in-depth virological analysis of a clinical cohort. *Microorganisms*. 2021;9:2145. doi:10.3390/microorganisms9102145
 110. Ip JD, Kok KH, Chan WM, et al. Intra-host non-synonymous diversity at a neutralizing antibody epitope of SARS-CoV-2 spike protein N-terminal domain. *Clin Microbiol Infect*. 2021;27:1350.e1-1350.e5. doi:10.1016/j.cmi.2020.10.030
 111. Siqueira JD, Goes LR, Alves BM, et al. SARS-CoV-2 genomic analyses in cancer patients reveal elevated intrahost genetic diversity. *Virus Evol*. 2022;7:veab013. doi:10.1093/ve/veab013
 112. Dudouet P, Colson P, Aherfi S, et al. SARS-CoV-2 quasi-species analysis from patients with persistent nasopharyngeal shedding. *Sci Rep*. 2022;12:18721. doi:10.1038/s41598-022-22060-z

113. Caccuri F, Messali S, Bortolotti D, et al. Competition for dominance within replicating quasispecies during prolonged SARS-CoV-2 infection in an immunocompromised host. *Virus Evol.* 2022;8:veac042. doi:10.1093/ve/veac042
114. Sonnleitner ST, Sonnleitner S, Hinterbichler E, et al. The mutational dynamics of the SARS-CoV-2 virus in serial passages in vitro. *Viol Sin.* 2022;37:198-207. doi:10.1016/j.virs.2022.01.029
115. Hodcroft EB. CoVariants: SARS-CoV-2 Mutations and Variants of Interest. 2021. <https://covariants.org/>
116. Peiffer-Smadja N, Bridier-Nahmias A, Ferré VM, et al. Emergence of E484K mutation following Bamlanivimab monotherapy among high-risk patients infected with the alpha variant of SARS-CoV-2. *Viruses.* 2021;13:1642. doi:10.3390/v13081642
117. Jensen B, Luebke N, Feldt T, et al. Emergence of the E484K mutation in SARS-COV-2-infected immunocompromised patients treated with bamlanivimab in Germany. *Lancet Reg Health Eur.* 2021;8:100164. doi:10.1016/j.lanep.2021.100164
118. Lohr B, Niemann D, Verheyen J. Bamlanivimab treatment leads to rapid selection of immune escape variant carrying the E484K mutation in a B.1.1.7-Infected and immunosuppressed patient. *Clin Infect Dis.* 2021;73:2144-2145. doi:10.1093/cid/ciab392
119. Guigon A, Faure E, Lemaire C, et al. Emergence of Q493R mutation in SARS-CoV-2 spike protein during bamlanivimab/etesevimab treatment and resistance to viral clearance. *J Infect.* 2022;84:248-288. doi:10.1016/j.jinf.2021.08.033
120. Vellas C, Del Bello A, Debarb A, et al. Influence of treatment with neutralizing monoclonal antibodies on the SARS-CoV-2 nasopharyngeal load and quasispecies. *Clin Microbiol Infect.* 2022;28:139.e5-139.e8. doi:10.1016/j.cmi.2021.09.008
121. Vellas C, Del Bello A, Gaube G, et al. Impact of casirivimab-imevumab on severe acute respiratory syndrome coronavirus 2 delta variant nasopharyngeal virus load and spike quasispecies. *Open Forum Infect Dis.* 2022;9:ofac093. doi:10.1093/ofid/ofac093
122. Jary A, Leducq V, Malet I, et al. Evolution of viral quasispecies during SARS-CoV-2 infection. *Clin Microbiol Infect.* 2020;26:1560.e1-1560.e4. doi:10.1016/j.cmi.2020.07.032
123. Bayarri-Olmos R, Rosbjerg A, Johnsen LB, et al. The SARS-CoV-2 Y453F mink variant displays a pronounced increase in ACE-2 affinity but does not challenge antibody neutralization. *J Biol Chem.* 2021;296:100536. doi:10.1016/j.jbc.2021.100536
124. Hartard C, Chaqroun A, Settembre N, et al. Multiorgan and vascular tropism of SARS-CoV-2. *Viruses.* 2022;14:515. doi:10.3390/v14030515
125. Gaiarsa S, Giardina F, Batisti Biffignandi G, et al. Comparative analysis of SARS-CoV-2 quasispecies in the upper and lower respiratory tract shows an ongoing evolution in the spike cleavage site. *Virus Res.* 2022;315:198786. doi:10.1016/j.virusres.2022.198786
126. Normandin E, Rudy M, Barkas N, et al. High-depth sequencing characterization of viral dynamics across tissues in fatal COVID-19 reveals compartmentalized infection. *Nat Commun.* 2023;14:574. doi:10.1038/s41467-022-34256-y
127. Adam G, Delbruck M. Reduction of dimensionality in biological diffusion processes. In: Rich A, Davidson N, eds. *Structural chemistry and molecular biology.* W.H. Freeman & Company; 1968:198-215.
128. Rosenberger CM, Brumell JH, Finlay BB. Microbial pathogenesis: lipid rafts as pathogen portals. *Curr Biol.* 2000;10:R823-R825. doi:10.1016/s0960-9822(00)00788-0
129. Fantini J, Azzaz F, Chahinian H, Yahi N. Electrostatic surface potential as a key parameter in virus transmission and evolution: how to manage future virus pandemics in the post-COVID-19 era. *Viruses.* 2023;15:284. doi:10.3390/v15020284
130. Devaux CA, Fantini J. ACE2 receptor polymorphism in humans and animals increases the risk of the emergence of SARSCoV-2 variants during repeated intra- and inter-species host-switching of the virus. *Front Microbiol.* 2023;14:1199561. doi:10.3389/fmicb.2023.1199561
131. Fantini J, Chahinian H, Yahi N. Convergent evolution dynamics of SARS-CoV-2 and HIV surface envelope glycoproteins driven by host cell surface receptors and lipid rafts: lessons for the future. *Int J Mol Sci.* 2023;24:1923. doi:10.3390/ijms24031923
132. Muller HJ. The relation of recombination to mutational advance. *Mutat Res.* 1964;106:2-9. doi:10.1016/0027-5107(64)90047-8
133. Colson P, Chaudet H, Delerce J, et al. Role of SARS-CoV-2 mutations in the evolution of the COVID-19 pandemic *bioRxiv.* Preprint posted online 2023. doi:10.1101/2023.05.01.538506
134. Koszalka P, George A, Dhanasekaran V, Hurt AC, Subbarao K. Effect of baloxavir and oseltamivir in combination on infection with influenza viruses with PA/I38T or PA/E23K substitutions in the ferret model. *mBio.* 2022;13:e0105622. doi:10.1128/mbio.01056-22
135. Codoñer FM, Darós JA, Solé RV, Elena SF. The fittest versus the flattest: experimental confirmation of the quasispecies effect with subviral pathogens. *PLoS Pathog.* 2006;2:e136. doi:10.1371/journal.ppat.0020136
136. Sun Y, Wang M, Lin W, Dong W, Xu J. Mutation blacklist" and "mutation whitelist" of SARS-CoV-2. *J Biosaf Biosecur.* 2022;4:114-120. doi:10.1016/j.job.2022.06.006
137. Benton MJ. The Red Queen and the Court Jester: species diversity and the role of biotic and abiotic factors through time. *Science.* 2009;323:728-732. doi:10.1126/science.1157719
138. Pecori R, Di Giorgio S, Paulo Lorenzo J, Nina Papavasiliou F. Functions and consequences of AID/APOBEC-mediated DNA and RNA deamination. *Nat Rev Genet.* 2022;23:505-518. doi:10.1038/s41576-022-00459-8
139. Jerzak GVS, Bernard K, Kramer LD, Shi PY, Ebel GD. The West Nile virus mutant spectrum is host-dependant and a determinant of mortality in mice. *Virology.* 2007;360:469-476. doi:10.1016/j.virol.2006.10.029
140. Sabrià A, Pintó RM, Bosch A, et al. Characterization of intra- and inter-host norovirus P2 genetic variability in linked individuals by amplicon sequencing. *PLoS One.* 2018;13:e0201850. doi:10.1371/journal.pone.0201850
141. Correa-Fiz F, Franzo G, Llorens A, et al. Porcine circovirus 2 (PCV2) population study in experimentally infected pigs developing PCV2-systemic disease or a subclinical infection. *Sci Rep.* 2020;10:17747. doi:10.1038/s41598-020-74627-3
142. Lagier JC, Armougom F, Million M, et al. Microbial culturomics: paradigm shift in the human gut microbiome study. *Clin Microbiol Infect.* 2012;18:1185-1193. doi:10.1111/1469-0691.12023
143. Kireev DE, Lopatukhin AE, Murzakova AV, et al. Evaluating the accuracy and sensitivity of detecting minority HIV-1 populations by Illumina next-generation sequencing. *J Virol Methods.* 2018;261:40-45. doi:10.1016/j.jviromet.2018.08.001
144. Stoler N, Nekrutenko A. Sequencing error profiles of Illumina sequencing instruments. *NAR Genomics Bioinformatics.* 2021;3:lqab019. doi:10.1093/nargab/lqab019
145. Karst SM, Ziels RM, Kirkegaard RH, et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with nanopore or PacBio sequencing. *Nat Methods.* 2021;18:165-169. doi:10.1038/s41592-020-01041-y
146. Riaz N, Leung P, Barton K, et al. Adaptation of Oxford nanopore technology for hepatitis C whole genome sequencing and identification of within-host viral variants. *BMC Genomics.* 2021;22:148. doi:10.1186/s12864-021-07460-1.
147. Yamauchi K, Sato M, Osawa L, et al. Analysis of direct-acting antiviral-resistant hepatitis C virus haplotype diversity by single-molecule and long-read sequencing. *Hepatal Commun.* 2022;6:1634-1651. doi:10.1002/hep4.1929
148. Charre C, Regue H, Dény P, et al. Improved hepatitis delta virus genome characterization by single molecule full-length genome

- sequencing combined with VIRIONT pipeline. *J Med Virol.* 2023;95:e28634. doi:10.1002/jmv.28634
149. Trémeaux P, Latour J, Ranger N, et al. SARS-CoV-2 co-infections and recombinations identified by long-read single-molecule real-time sequencing. *Microbiol Spectr.* 2023;11:e0049323. doi:10.1128/spectrum.00493-23
150. Ng TTL, Su J, Lao HY, et al. Long-read sequencing with hierarchical clustering for antiretroviral resistance profiling of mixed human immunodeficiency virus quasispecies. *Clin Chem.* 2023;69:1174-1185. doi:10.1093/clinchem/hvad108
151. Bradley CC, Gordon AJE, Wang C, et al. RNA polymerase inaccuracy underlies SARS-CoV-2 variants and vaccine heterogeneity. *Res Sq.* 2022. doi:10.21203/rs.3.rs-1690086/v1

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Colson P, Bader W, Fantini J, et al. From viral democratic genomes to viral wild bunch of quasispecies. *J Med Virol.* 2023;95:e29209. doi:10.1002/jmv.29209