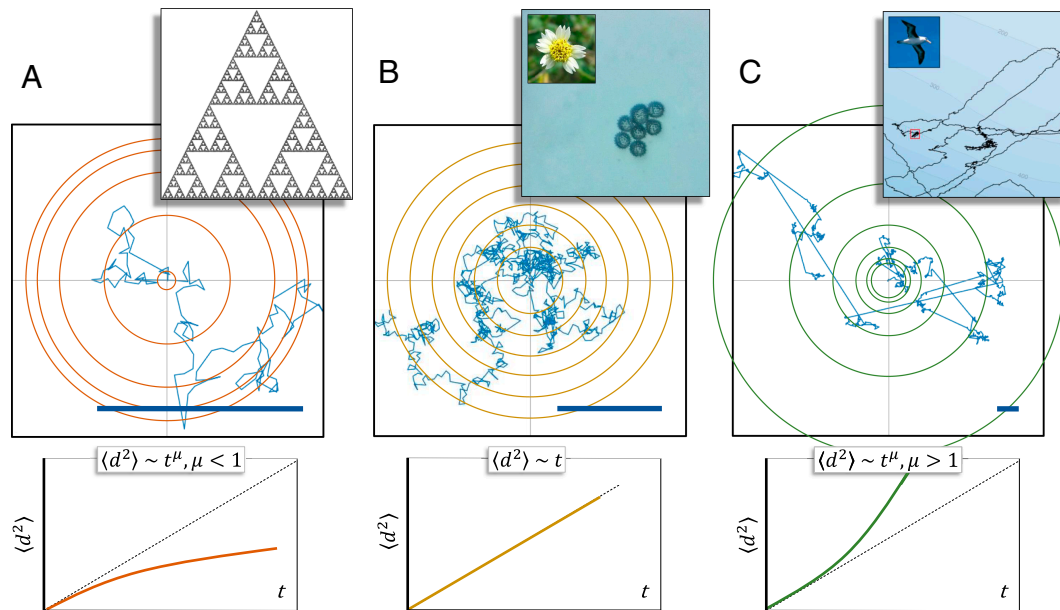# Physics of diffusion in viral genome evolution

Susanna Manrubia[a,b,1] 🆔 and José A. Cuesta[b,c,d] 🆔

Fig. 1. Normal and anomalous stochastic diffusion. Various features of (*A*) subdiffusion, (*B*) normal diffusion, and (*C*) superdiffusion are depicted. In the three large mid panels, for each diffusive class, the trajectory of a particle is represented for the same diffusing time, highlighting quantitative differences in the typical distance covered. For comparison, the horizontal blue bar represents the same distance in every panel, while concentric circles schematically stand for the distance explored after equal time increments (from one circle to the next). *Bottom* panels represent, for each diffusive class, the growth of the MSD $\langle d^2 \rangle$ with time t. (*A*) Subdiffusion: MSD grows slower than time, such that reaching long distances takes progressively longer times. An example of subdiffusion is the movement of a particle randomly jumping between adjacent sites of a fractal structure. Sierpinski's gasket (*Upper* panel), for example, would trap the particle in dense regions and cause arbitrarily long delays, responsible for subdiffuse behavior. Primal, Alpha, and Omicron variants follow subdiffusive dynamics in the space of sequences (1). (*B*) Normal diffusion: this is the phenomenon described by Einstein's theory and first observed by Brown. The MSD grows proportional to time. The top panel represents pollen grains of Tridax and the flower. Source: https://commons.wikimedia.org/wiki/File:The_pollen_grains_of_Tridax.jpg and https://commons.wikimedia.org/wiki/File:Tridax_procumbens_flower.jpg; (*C*) Superdiffusion: the particle undergoes long directed jumps that cause the MSD to grow fast with time. This movement is observed in the foraging behavior of some animal species, as in the case of the black-browed albatross—an example of whose foraging trajectories are represented in the *Top* panel. Reprinted from ref. 10. Delta variant of SARS-CoV-2 spreads superdiffusively.

Using data from SARS-CoV-2 viral genomes sequenced during the COVID-19 pandemic, Goiriz and colleagues analyze, in PNAS (1), the way that mutational variants "move" through the space of sequences of this virus. Their analysis reveals an unexpected phenomenon known as "anomalous diffusion" in SARS-CoV-2 genomes. This means that instead of undergoing an exploration of neighboring variants through replication and unconstrained mutation (akin to normal diffusion), the virus exhibits either a hindered exploration, progressing at a slower-than-expected speed (subdiffusive spread), or an accelerated exploration, displaying a faster-than-expected spread through the sequence space (superdiffusive spread).

In order to understand the deep implications of this finding, we must travel back 200 y in time. Diffusion, a well-established physical phenomenon since the early 19th century, describes the intermixing of gases or liquids. A classic example is the diffusion of an ink droplet in a glass of water. Mathematically, diffusion was described by Fick in 1855 (2) through an equation known as the diffusion equation or Fick's law, which Fick derived by drawing from an analogy to

heat conduction. However, the microscopic explanation of this phenomenon was hidden in a perplexing observation coming from an entirely different scientific discipline.

Robert Brown was a renowned botanist fascinated with the mechanisms of fertilization in flowering plants. In June 1827, while studying pollen particles of *Clarkia pulchella* suspended in water, he witnessed the incessant agitation

Author affiliations: [a]Departamento de Biología de Sistemas, Centro Nacional de Biotecnología (CSIC), 28049 Madrid, Spain; [b]Grupo Interdisciplinar de Sistemas Complejos, 28911 Madrid, Spain; [c]Instituto de Biocomputación y Física de Sistemas Complejos, Campus Río Ebro, Universidad de Zaragoza, 50018 Zaragoza, Spain; and [d]Departamento de Matemáticas, Universidad Carlos III de Madrid, 28911 Leganés, Spain

of these particles. His meticulous observations were supported by a pioneering and proficient use of a specially designed microscope for investigating "minute points" (3). The irregular motion he described came to be known as "Brownian motion" and became inseparably associated with his name. Brown, however, could not discern the causes underlying the behavior of those seemingly inanimate, but active, particles. It was not until 1905, when Einstein published his work on the theory of Brownian motion, that an explanation emerged (4). Einstein correctly attributed the motion of the particles to the thermal agitation of water molecules, which would push the pollen granules in random directions after each collision. On a larger scale, it is the collective behavior of numerous minute particles that causes the spread of pollen grains on water. Einstein formulation had quantifiable consequences: If a large number of tiny particles are initially positioned in a specific location, the mean square distance covered by these particles after a time t follows the simple law $\langle d^2 \rangle = 6Dt$, where D represents the diffusion constant appearing in Fick's law. Incidentally, Einstein's explanation of Brownian motion was considered the definitive proof of the existence of atoms—by the time a controversial hypothesis.

> **Goiriz et al.'s study introduces a significant novelty by characterizing the dynamics of how the population of SARS-CoV-2 genomes explores its mutational neighborhood within each variant.**

Today, Brownian motion (or random walk) encompasses a family of stochastic processes in which a "particle" undergoes random jumps in time and space, a way of modeling the unpredictable nature of molecule collisions. The long-term behavior of its mean square displacement (MSD) follows the law $\langle d^2 \rangle \sim t^\mu$. Ordinary Brownian motion (diffusion) is characterized by $\mu = 1$, as elucidated by Einstein. Any deviation from this exponent indicates anomalous diffusion, with $\mu > 1$ signifying superdiffusion and $\mu < 1$ indicating subdiffusion (Fig. 1).

Multiple models of Brownian motion have been proposed to uncover the underlying causes of anomalous diffusion. These models incorporate two key elements: the waiting time distribution of the particle at a given position and the jump-length distribution. Ordinary diffusion occurs when both distributions have finite variances. Failure to meet either condition leads to anomalous diffusion (5). A long-tailed waiting time distribution hampers the particle's movement, resulting in subdiffusion (Fig. 1A). The disordered nature of the medium is a prominent factor contributing to this hindered dynamics (6, 7). Examples of subdiffusive dynamics include the motion of mRNA molecules inside living cells (8) or the search for a target of DNA-binding proteins in mammalian cells (9). When the finite variance condition is violated by the jump-length distribution, the particle occasionally makes long jumps (Fig. 1C), leading to a dynamics commonly referred to as Lévy flights or Lévy walks, which exhibits superdiffusion (5). Typical examples of these Lévy dynamics are observed in foraging behaviors (10, 11).

Goiriz et al. (1) conducted an extensive analysis using a dataset comprising over 2.7 million publicly available SARS-CoV-2 genomes obtained from the GISAID database (https://www.gisaid.org). The analyzed genomes were collected in the United Kingdom and came with annotated information, including the acquisition time of each sequence and its corresponding strain. The authors focused on analyzing five different major groups: Primal (early sequences not associated with a particular lineage), Alpha, Delta, and two distinct Omicron strains. To capture the temporal dynamics, the genomes were classified with a resolution of one week, enabling the evaluation of diversity at fixed moments and tracking its variations over time. The authors annotated the number and position of different mutations present in these sequences. The main objective of their study was to characterize the generation of genomic diversity during the evolution of SARS-CoV-2 and to identify any differences between the considered strains and the average behavior of the virus.

A notable finding is that the incorporation of mutations into the SARS-CoV-2 genome does not follow a constant rate over time. Instead, there are periods of slow accumulation interspersed with sudden increases. This indicates that the molecular clock governing mutation rates varies over time and, as documented by the researchers, the rate depends on the specific variant being considered. On average, SARS-CoV-2 acquires one new mutation every 11 d. However, for the Primal, Alpha, and Delta variants, it took 19 d for a single mutation to become fixed. In the case of the Omicron strain, this figure may rise to 32 d. While these rates are comparatively slower, they are compensated by an accelerated fixation of mutations when variants replace each other. For instance, there was a burst of mutation accumulation when the Alpha variant replaced the Primal variant or when the Omicron variant replaced the Delta variant as the dominant circulating strain. These bursts coincided with an increase in both viral genomic diversity, associated with an overdispersion of the molecular clock (12, 13), and the number of infections. This last observation resembles punctuated equilibrium at the molecular level, a phenomenon that was numerically predicted (14), characterized in various synthetic evolutionary models (15), and empirically observed in influenza A as well (16, 17).

But Goiriz et al.'s study introduces a significant novelty by characterizing the dynamics of how the population of SARS-CoV-2 genomes explores its mutational neighborhood within each variant. By following the sequences in their dataset, we can contemplate the movement of the virus over weeks, similar to an ensemble of random walkers wandering around physical space. In a fully neutral fitness landscape, where all mutants are equally accessible and have the same selective value, normal diffusion is to be expected. To quantify the diffusive behavior, Goiriz et al. plotted the relation between the MSD of genomes at time t—with distance evaluated as the number of mutations a sequence had with respect to a reference sequence—as a function of t, as in the bottom plots of Fig. 1. They found that most variants struggle to explore genomes that are

only a few mutations away and remain close in sequence space for a longer duration than expected, as it happens with Primal, Alpha, and Omicron, which exhibit subdiffusive dynamics.

Genomic sequences can be conceptualized as nodes on a high-dimensional lattice. In this framework, two sequences are considered neighbors if they differ at a single position, reflecting the mutational dynamics. Consequently, the exploration of sequence space can be visualized as the "diffusion" of hypothetical particles in this high-dimensional space, with the distance measured by the number of positions (mutations) that separate two sequences. However, the exploration of sequence space is affected by the existence of a genotype-to-phenotype (GP) map, which assigns genomic sequences to expressed phenotypes (18). In both realistic models and natural systems, this relationship is highly redundant, as many genotypes typically map to the same phenotype. This redundancy allows for the exploration of genotype space and the localization of numerous alternative phenotypes without losing functionality (19, 20). However, the structure of phenotype networks is not uniform, meaning that genotypes in the network exhibit significant variation in the number of neutral neighbors they possess (21). Consequently, the drifting caused by point mutations on a phenotype network is influenced by the presence of an underlying, heterogeneous network structure, which can lead to deviations from unconstrained, normal diffusion. With this understanding, one could argue that it is the disorder induced by the GP map that results in the local trapping of genomes associated with a variant, providing a qualitative explanation for why most variants exhibit subdiffusive exploration of genomes. However, Delta experiences superdiffusive propagation, a phenomenon that is difficult to explain in light of our current knowledge.

The process of evolution and adaptation encompasses a multitude of distinct processes and constraints, both intrinsic and extrinsic, whose influence on the final outcome remains a significant open question. Evolution and adaptation on complex, high-dimensional fitness landscapes occur in a discontinuous manner. Periods of search, influenced by intricate mappings between sequences and functional organisms (22), are intermittently interrupted by sudden jumps in genotype space due to factors such as the selection of phenotypes with higher adaptive value, environmental changes, or a combination of both. As our understanding of the molecular dynamics of populations deepens, classical expectations regarding the characteristics of the evolutionary process are being challenged. The generation of diversity is not uniformly distributed across time and genome spaces; mutations accumulate irregularly and, more often than not, in an unpredictable manner. In all likelihood, it will be revealed that anomalous diffusion is not an extravagant property of searches in sequence spaces.

Goiriz et al. (1) have provided compelling evidence of this behavior in the context of SARS-CoV-2, prompting the need to investigate comparable datasets for other viruses and potentially even cellular organisms. Although two centuries separate their observations from Brown's description of pollen grain movement, both discoveries share parallelisms that are inherent to scientific research: surprise precedes understanding. The diffusion of particles in stationary liquids and various other diffusive processes is now well understood. A comprehensive understanding of the fundamental principles underlying the anomalous diffusion of virus genomes in sequence space is still awaited.

1. L. Goiriz, R. Ruiz, O. Garibo-i-Orts, J. A. Conejero, G. Rodrigo, A variant-dependent molecular clock with anomalous diffusion models SARS-CoV-2 evolution in humans. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2303578120 (2023).
2. A. Fick, Über diffusion. *Ann. Phys.* **170**, 59–86 (1855).
3. R. Brown, A brief account of microscopical observations on the particles contained in the pollen of plants and the general existence of active molecules in organic and inorganic bodies. *Philos. Mag.* **4**, 161–173 (1828).
4. A. Einstein, Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Ann. Phys.* **322**, 549–560 (1905).
5. R. Metzler, J. Klafter, The random walk's guide to anomalous diffusion: A fractional dynamic approach. *Phys. Rep.* **339**, 1–77 (2000).
6. J. P. Bouchaud, A. Georges, Anomalous diffusion in disordered media: Statistical mechanisms, models and physical applications. *Phys. Rep.* **195**, 127–293 (1990).
7. S. Havlin, D. Ben-Avraham, Diffusion in disordered media. *Adv. Phys.* **51**, 187–292 (2002).
8. I. Golding, E. C. Cox, Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.* **96**, 098102 (2006).
9. D. Normanno *et al.*, Probing the target search of DNA-binding proteins in mammalian cells using TetR as model searcher. *Nat. Commun.* **6**, 7357 (2015).
10. N. E. Humphries, H. Weimerskirch, N. Queiroz, E. J. Southall, D. W. Sims, Foraging success of biological Lévy flights recorded in situ. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 7169–7174 (2012).
11. R. Nathan *et al.*, Big-data approaches lead to an increased understanding of the ecology of animal movement. *Science* **375**, eabg1780 (2022).
12. A. Raval, Molecular clock on a neutral network. *Phys. Rev. Lett.* **99**, 138104 (2007).
13. S. Manrubia, J. A. Cuesta, Evolution on genotype networks accelerates the ticking rate of the molecular clock. *J. R. Soc. Interface* **12**, 20141010 (2015).
14. M. A. Huynen, P. F. Stadler, W. Fontana, Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 397–401 (1996).
15. J. Aguirre, P. Catalán, J. A. Cuesta, S. Manrubia, On the networked architecture of genotype spaces and its critical effects on molecular evolution. *Open Biol.* **8**, 180069 (2018).
16. K. Koelle, S. Cobey, B. Grenfell, M. Pascual, Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**, 1898–1903 (2006).
17. E. van Nimwegen, Influenza escapes immunity along neutral networks. *Science* **314**, 1884–1886 (2006).
18. P. F. Stadler, B. M. R. Stadler, Genotype-phenotype maps. *Biol. Theor.* **1**, 268–279 (2006).
19. J. Maynard Smith, Natural selection and the concept of a protein space. *Nature* **225**, 563–564 (1970).
20. J. Aguilar-Rodríguez, J. L. Payne, A. Wagner, A thousand empirical adaptive landscapes and their navigability. *Nat. Ecol. Evol.* **1**, 45 (2017).
21. S. F. Greenbury, S. Schaper, S. E. Ahnert, A. A. Louis, Genetic correlations greatly increase mutational robustness and can both reduce and enhance evolvability. *PLoS Comput. Biol.* **12**, e1004773 (2016).
22. S. Manrubia *et al.*, From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Phys. Life Rev.* **38**, 55–106 (2021).