

PEARLS

Monitoring and redirecting virus evolution

Gonzalo Moratorio, Marco Vignuzzi*

Viral Populations and Pathogenesis Unit, Institut Pasteur, CNRS UMR 3569, Paris, France

* marco.vignuzzi@pasteur.fr

Introduction

“Nothing in biology makes sense except in the light of evolution.”—Theodosius Dobzhansky

Viral pandemics can kill more people than war. From biblical plagues to modern-day outbreaks, the ensuing devastation is in part driven by the extreme genetic diversity of these pathogens, which allows them to rapidly evolve to escape immunity, jump to new species, and enter new ecological niches. The quote above, penned by a central figure who helped shape the modern-day synthesis of evolutionary science, is oft encountered today as a reminder that evolution remains relevant to all biology and certainly to infectious diseases. Indeed, throughout our history, whether deliberately or unintentionally, we humans have tried to control this very force responsible for our existence. And more recently, as microbiologists, advances in computer science and sequencing technology have moved our field to the next level. We can now follow, almost in real time, the molecular epidemiology of emerging outbreaks, allowing us to directly test field observations in the lab. What methods are available to follow viral evolution? Is it possible to predict virus evolution? Can we use this knowledge to drive deadly pathogens towards evolution’s dead end, to extinction? These questions will guide the reader through this review.

How to first characterize the emergence of viral outbreaks? By sequencing and phylogenetics

We employ phylogenetic approaches to study the evolutionary history and relationships of organisms. Heritable traits, such as genetic sequences, are represented in phylogenetic trees (Fig 1, top panel, center), which portray relatedness between organisms by branches and in which the length of branching lines represent evolutionary time. Phylogenetics provides an integrative perspective of the identity, classification, ecology, and evolutionary history of viral strains [1]. Usually, it involves mathematical models implementing mainly three kinds of methods: a) parsimony, a criterion of ‘frugality’, based on favouring the simplest explanation or fit of the data, in which the best phylogenetic tree is one that requires the fewest evolutionary changes; b) maximum likelihood (ML), based on inferring probability distributions such that the more likely a viral strain (sequence) falls within a cluster of a given tree, the more that tree is validated; and c) Bayesian inference, which relies on ML but integrates the probability of hypothesis according to previous data (prior probability) to define the best tree. All aforementioned methods are used to analyze genetic relationships among viral strains [2]. Bayesian approximations favoured the integration of sophisticated sampling methods such as Markov chain Monte Carlo (MCMC) algorithms, revolutionizing phylogenetics through the incorporation of complex models of evolution and the estimation of parameters such as substitution



OPEN ACCESS

Citation: Moratorio G, Vignuzzi M (2018) Monitoring and redirecting virus evolution. *PLoS Pathog* 14(6): e1006979. <https://doi.org/10.1371/journal.ppat.1006979>

Editor: Matthew J. Evans, Mount Sinai School of Medicine, UNITED STATES

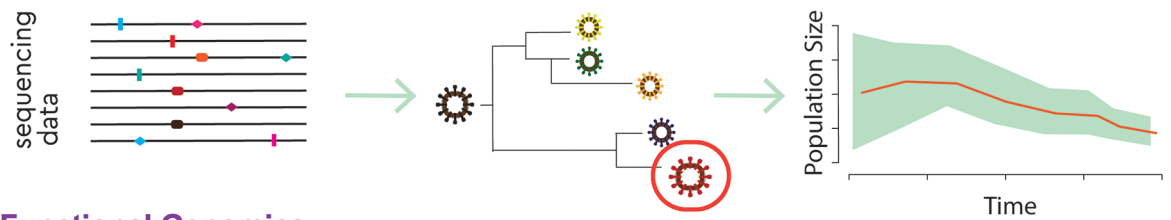
Published: June 7, 2018

Copyright: © 2018 Moratorio, Vignuzzi. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

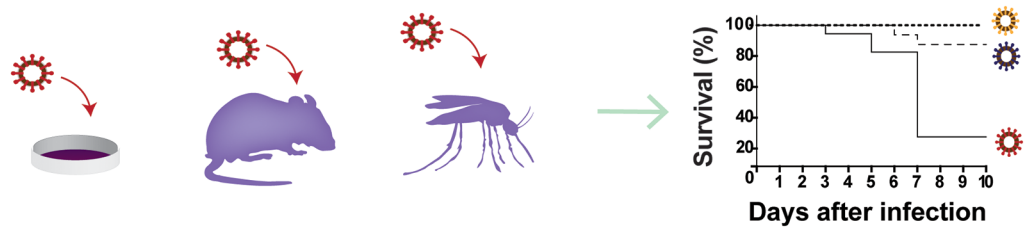
Funding: This work was supported by the Laboratoire d’Excellence “Integrative Biology of Emerging Infectious Diseases” (grant ANR-10-LABX-62-IBEID) and the Equipe FRM DEQ20150331759 grant from the French Fondation pour la Recherche Médicale (to MV). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

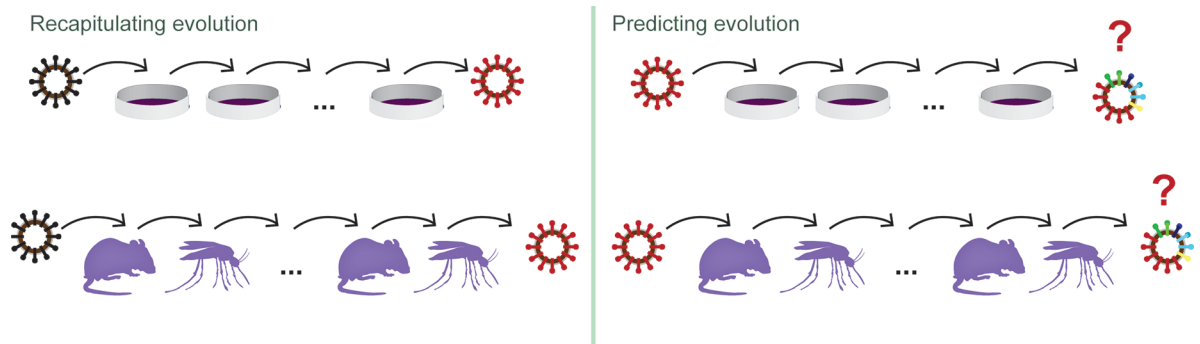
Phylogenetic approaches



Functional Genomics



Experimental Evolution



Fitness Landscape representation

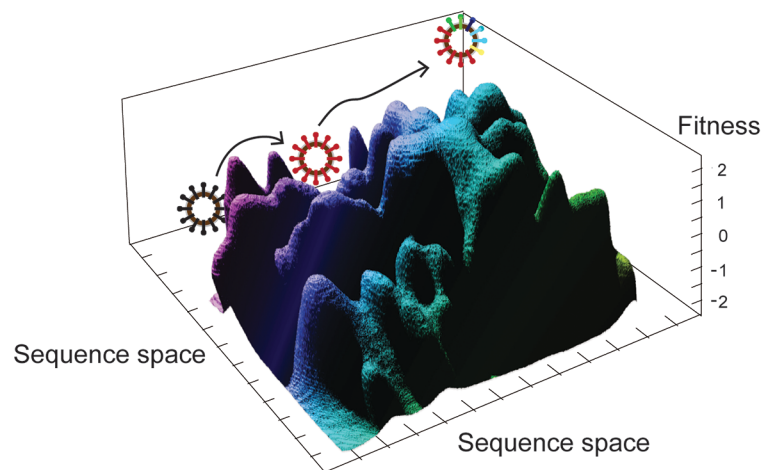


Fig 1. Monitoring virus evolution. Top panel: Phylogenetic approaches are first used to characterize a viral outbreak. From left to right: Sequencing data identifies a new viral genotype by phylogenetics. This new strain is represented in red in the phylogenetic tree. By combining these data with sampling dates, a Bayesian skyline plot reveals the demographic history of the epidemic. Middle panels: First, by

functional genomics, this new variant is tested *in vitro* and *in vivo*. Survival curves show the phenotype of this new strain. Second, experimental evolution is performed to a) recapitulate the evolution observed in nature using ancestral genotypes and b) predict the next mutations likely to emerge, using as a starting point the newly identified variant (in red). Bottom panel: Implementation of genotype-phenotype maps would help monitor evolution and potentially predict future trajectories towards, or away from, virulence.

<https://doi.org/10.1371/journal.ppat.1006979.g001>

rates, divergence times, and other population genetics patterns [3]. In addition, the number of infections over time, inferred from Bayesian skyline plots, allows the reconstruction of the demographic history of a pathogen during epidemics (Fig 1, top panel, right) [4–6]. These advances were significantly boosted by the increasingly abundant genomic data from next-generation sequencing techniques, which are moving from core facilities to benchtop and even to field-friendly platforms. Altogether, this workflow makes near-real-time molecular epidemiology a feasible endeavour. These approaches were used to reconstruct the emergence and spread of recent Zika virus epidemics in the Americas, identifying the timing and sources of introductions in different geographical regions [7]. Similarly, real-time molecular epidemiology of the recent Ebola outbreak in West Africa revealed considerable disconnection between transmission clusters and independently evolved lineages [8,9]. Phylogenetics thus builds a clearer picture of virus outbreaks by identifying evolutionary and transmission patterns.

How to test new findings made from field samples? By comparing the phenotype of new mutations with the original strain

After phylogenetic characterization, newly identified mutations or viral strains can be tested in the lab in what is known as functional genomics. These approaches, which attempt to mimic what is happening in the wild, reveal how new mutations impact viral phenotype. There are many examples of newly identified mutations associated with antiviral resistance, higher pathogenicity, increased transmission, and cross-species transmission that were confirmed by these means [10–13]. For instance, a single mutation (A188V) in a nonstructural protein (NS1) of Zika virus was first identified by phylogenetics and then shown to increase infectivity in laboratory mosquitoes [14]. The E1-A226V mutation in chikungunya virus that contributed to its jump from *Aedes aegypti* to *A. albopictus* mosquitoes was also identified by phylogenetics and later confirmed by functional genomics [15,16]. Sequence analysis of viruses from an outbreak of highly pathogenic H7N7 avian flu in the Netherlands identified 15 amino acid differences between a fatal case and other isolates [17]. Functional genomic studies in mice later identified a previously reported host range mutation (E627K) in the PB2 gene [18,19] that increased viral pathogenesis *in vivo* [20]. On the other hand, new phenotypes can also result from a constellation of mutations at the viral population level. Although these observations have been only seen in tissue culture, they are still significant and could be the starting point to test it *in vivo* [21,22]. However, to better achieve mechanistic understandings, we still need to focus on a few mutations and their immediate effect on fitness. Thus, once a mutation of interest is identified and its impact on phenotype confirmed, one can set out to determine under what conditions this emergence event occurred. To do so, experimental evolution can be performed *in vitro* or *in vivo*.

How to reproduce in the lab the genetic and phenotypic changes that have occurred in nature? By experimental evolution

Experimental evolution tries to recreate the evolutionary dynamics surrounding the emergence of a new adaptive mutation in a controlled environment and within a specific time frame. By testing different host environments or viral genotypes, one tries to determine why some mutations occur and others do not. In our own lab, we showed that the chikungunya

virus E1-A226V adaptive mutation, mentioned above, arises readily in *A. albopictus* mosquitoes in only seven days, and we identified potentially newly emerging mutations in the same gene [23]. In studying the emergence of canine parvovirus by experimental evolution in vitro, Allison and colleagues recapitulated several mutations observed in natural isolates that were specific to host and virus genetic background and confirmed their fitness advantages over the parental virus [24]. Experimentation can be taken one step further, to identify the determinants that can either increase transmissibility, as was performed for H5N1 influenza virus [25], or decrease virulence, which has been used in live-vaccine attenuation for decades [26]. These approaches thus require special consideration before, during, and after the experiments, since experimental evolution in specialized environments could lead to gain-of-function mutations. While powerful in approach, experimental evolution does have limitations, including stochasticity, population size, limited environments, and resources, and although we cannot be sure these mutations will appear in nature, their identification does give us an advantage as chess masters by providing a short list of ‘next moves’ available to a virus.

How to determine whether new viral genotypes can lead to more pathogenic phenotypes? By implementing fitness landscapes

The fitness landscape is a helpful concept to metaphorically visualize evolution and to uncover how genotype relates to phenotype, a central goal in evolutionary biology [27]. We can consider fitness landscapes as a sort of Global Positioning System (GPS) device, whereby knowing the coordinates of an individual (genotype) allows us to predict how well an organism is positioned in its current environment (fitness, phenotype) and where it is likely to go next. Usually, fitness is represented by the “height” of the landscape, which is grounded on a plane that denotes the genotypic space: Peaks are good places to be (higher fitness), whereas valleys are not (lower fitness). This metaphor has been inching towards reality thanks to multidisciplinary approaches that combine computer science and applied maths with high throughput, highly quantitative biological data [28]. For instance, more than 70,000 HIV-1 samples were used to measure the reverse transcriptase and protease fitness in the presence and absence of different antiviral drugs. Fitness measurements were then coupled with sequencing data and fitted to different mathematical models. This work provided a predictive model based on a biologically relevant fitness landscape, which predicted more than 50% of the fitness values obtained [29]. On the other hand, the process of acquiring beneficial mutations could thus be visualized within the fitness landscape. Such adaptive walks could be further investigated by reconstructing ancestral genotypes and submitting them to experimental evolution. Indeed, this was the case in recapitulating the evolutionary pathway to virulence of the oral polio vaccine [30]. Knowing the ancestral virus and sequences from vaccine-derived poliovirus infections from different outbreaks, Stern and colleagues combined mathematical models and experimental evolution to retrace each mutation and recombination back to the virulent form [30]. Although the trade-off between fitness and pathogenesis is not always reciprocal, a better understanding of genotype–phenotype maps would help monitor evolution and potentially predict future trajectories towards virulence.

Can we modify an RNA virus’s evolution to our own benefit?

As previously mentioned, the valleys in fitness landscapes are detrimental regions, corresponding to mutations on nonviable or low-fitness genotypes. In principle, we could use this knowledge to redirect virus evolution towards a more dismal future by increasing the likelihood of acquiring such detrimental mutations. Among substitutions, stop codons are possibly the worst mutation an organism could acquire. This idea was directly tested in our recent work in

which two very different RNA viruses, influenza A and Coxsackie B3, were genetically engineered to generate more stop codons after replication occurs. These viruses were re-coded in serine and leucine codons by synonymous mutations that placed them only one mutation away from stop codons. Both engineered viruses generated more stop mutations, *in vitro* and *in vivo*, accompanied by significant losses in viral fitness and attenuation in mouse models [31]. These findings were in agreement with previous studies in which the ability of RNA viruses to buffer mutation (known as genetic robustness) was reduced. Importantly, these works show that the same error rate will differently impact a genome depending on the codons it carries, which can lead to distinct amino acid changes. Thus, the position that a virus occupies in genotypic space defines not only its mutational robustness but also its mutant spectrum, the evolutionary trajectories available to it, and ultimately its evolvability [32]. The endeavours described are still in their infancy, but the goal to characterize the local sequence and fitness landscape of a virus, to monitor and potentially predict its immediate future, and to evaluate how antiviral approaches might alter these trajectories seems attainable. The further integration of mathematical modelling, bioinformatics, and experimental evolution will booster our response to viral outbreaks.

Acknowledgments

We thank Nathan Grubaugh, Alvaro Fajardo, Enzo Poirier, Stephanie Beaucourt, and Lucía Carrau for helpful discussions.

References

1. Holmes EC. The evolution and emergence of RNA viruses. Oxford University Press; 2009.
2. Yang Z, Rannala B. Molecular phylogenetics: principles and practice. Nat Rev Genet. Nature Publishing Group; 2012; 13: 303–314. <https://doi.org/10.1038/nrg3186> PMID: 22456349
3. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP. Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. Science. 2001; 294: 2310–2314. <https://doi.org/10.1126/science.1065889> PMID: 11743192
4. Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. Bioinforma Oxf Engl. 2001; 17: 754–5.
5. Lemey P, Rambaut A, Drummond AJ, Suchard MA, Ali Y. Bayesian Phylogeography Finds Its Roots. Fraser C, editor. PLoS Comput Biol. Academic Press; 2009; 5: e1000520. <https://doi.org/10.1371/journal.pcbi.1000520> PMID: 19779555
6. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. BioMed Central; 2007; 7: 214. <https://doi.org/10.1186/1471-2148-7-214> PMID: 17996036
7. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature. 2017; 546: 401–405. <https://doi.org/10.1038/nature22400> PMID: 28538723
8. Dudas G, Carvalho LM, Bedford T, Tatem AJ, Baele G, Faria NR, et al. Virus genomes reveal factors that spread and sustained the Ebola epidemic. Nature. 2017; 544: 309–315. <https://doi.org/10.1038/nature22040> PMID: 28405027
9. Carroll MW, Matthews DA, Hiscox JA, Elmore MJ, Pollakis G, Rambaut A, et al. Temporal and spatial analysis of the 2014–2015 Ebola virus outbreak in West Africa. Nature. 2015; 524: 97–101. <https://doi.org/10.1038/nature14594> PMID: 26083749
10. Schountz T, Baker ML, Butler J, Munster V. Immunological Control of Viral Infections in Bats and the Emergence of Viruses Highly Pathogenic to Humans. Front Immunol. 2017; 8: 1098. <https://doi.org/10.3389/fimmu.2017.01098> PMID: 28959255
11. Parrish CR, Holmes EC, Morens DM, Park E-C, Burke DS, Calisher CH, et al. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev MMBR. 2008; 72: 457–470. <https://doi.org/10.1128/MMBR.00004-08> PMID: 18772285
12. Tsetsarkin KA, Vanlandingham DL, McGee CE, Higgs S. A Single Mutation in Chikungunya Virus Affects Vector Specificity and Epidemic Potential. PLoS Pathog. 2007; 3: e201. <https://doi.org/10.1371/journal.ppat.0030201> PMID: 18069894

13. Yuan L, Huang X-Y, Liu Z-Y, Zhang F, Zhu X-L, Yu J-Y, et al. A single mutation in the prM protein of Zika virus contributes to fetal microcephaly. *Science*. 2017; eaam7120. <https://doi.org/10.1126/science.aam7120> PMID: 28971967
14. Liu Y, Liu J, Du S, Shan C, Nie K, Zhang R, et al. Evolutionary enhancement of Zika virus infectivity in *Aedes aegypti* mosquitoes. *Nature*. 2017;545: nature22365. <https://doi.org/10.1038/nature22365> PMID: 28514450
15. Schuffenecker I, Itman I, Michault A, Murri S, Frangeul L, Vaney M-C, et al. Genome Microevolution of Chikungunya Viruses Causing the Indian Ocean Outbreak. *PLoS Med*. 2006; 3: e263. <https://doi.org/10.1371/journal.pmed.0030263> PMID: 16700631
16. Tssetsarkin KA, Weaver SC. Sequential Adaptive Mutations Enhance Efficient Vector Switching by Chikungunya Virus and Its Epidemic Emergence. *PLoS Pathog*. 2011; 7: e1002412. <https://doi.org/10.1371/journal.ppat.1002412> PMID: 22174678
17. Fouchier RAM, Schneeberger PM, Rozendaal FW, Broekman JM, Kemink SAG, Munster V, et al. Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc Natl Acad Sci U S A*. 2004; 101: 1356–1361. <https://doi.org/10.1073/pnas.0308352100> PMID: 14745020
18. Subbarao EK, London W, Murphy BR. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. *J Virol*. 1993; 67: 1761–1764. PMID: 8445709
19. Gabriel G, Dauber B, Wolff T, Planz O, Klenk H-D, Stech J. The viral polymerase mediates adaptation of an avian influenza virus to a mammalian host. *Proc Natl Acad Sci*. 2005; 102: 18590–18595. <https://doi.org/10.1073/pnas.0507415102> PMID: 16339318
20. Munster VJ, de Wit E, van Riel D, Beyer WEP, Rimmelzwaan GF, Osterhaus ADME, et al. The molecular basis of the pathogenicity of the Dutch highly pathogenic human influenza A H7N7 viruses. *J Infect Dis*. 2007; 196: 258–265. <https://doi.org/10.1086/518792> PMID: 17570113
21. Bordería AV, Isakov O, Moratorio G, Henningsson R, Agüera-González S, Organtini L, et al. Group Selection and Contribution of Minority Variants during Virus Adaptation Determines Virus Fitness and Phenotype. *PLoS Pathog*. 2015; 11: e1004838. <https://doi.org/10.1371/journal.ppat.1004838> PMID: 25941809
22. Xue KS, Hooper KA, Ollodart AR, Dingens AS, Bloom JD. Cooperation between distinct viral variants promotes growth of H3N2 influenza in cell culture. *eLife*. 2016; 5: e13974. <https://doi.org/10.7554/eLife.13974> PMID: 26978794
23. Stapleford KA, Coffey LL, Lay S, Bordería AV, Duong V, Isakov O, et al. Emergence and transmission of arbovirus evolutionary intermediates with epidemic potential. *Cell Host Microbe*. 2014; 15: 706–716. <https://doi.org/10.1016/j.chom.2014.05.008> PMID: 24922573
24. Allison AB, Kohler DJ, Ortega A, Hoover EA, Grove DM, Holmes EC, et al. Host-Specific Parvovirus Evolution in Nature Is Recapitulated by In Vitro Adaptation to Different Carnivore Species. *PLoS Pathog*. 2014; 10: e1004475. <https://doi.org/10.1371/journal.ppat.1004475> PMID: 25375184
25. Herfst S, Schrauwen EJA, Linster M, Chutinimitkul S, de Wit E, Munster VJ, et al. Airborne transmission of influenza A/H5N1 virus between ferrets. *Science*. 2012; 336: 1534–1541. <https://doi.org/10.1126/science.1213362> PMID: 22723413
26. Plotkin SA, Plotkin SL. The development of vaccines: how the past led to the future. *Nat Rev Microbiol*. 2011; 9: 889. <https://doi.org/10.1038/nrmicro2668> PMID: 21963800
27. Wright S. The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress of Genetics*; 1932; Ithaca, New York, USA. Available from: <http://www.esp.org/books/6th-congress/facsimile/contents/6th-cong-p356-wright.pdf>. Accessed 5/21/18.
28. de Visser JAGM, Krug J. Empirical fitness landscapes and the predictability of evolution. *Nat Rev Genet*. 2014; 15: 480–490. <https://doi.org/10.1038/nrg3744> PMID: 24913663
29. Hinkley T, Martins J, Chappey C, Haddad M, Stawiski E, Whitcomb JM, et al. A systems analysis of mutational effects in HIV-1 protease and reverse transcriptase. *Nat Genet*. 2011; 43: 487–489. <https://doi.org/10.1038/ng.795> PMID: 21441930
30. Stern A, Yeh MT, Zinger T, Smith M, Wright C, Ling G, et al. The Evolutionary Pathway to Virulence of an RNA Virus. *Cell*. 2017; 169: 35–46.e19. <https://doi.org/10.1016/j.cell.2017.03.013> PMID: 28340348
31. Moratorio G, Henningsson R, Barbezange C, Carrau L, Bordería AV, Blanc H, et al. Attenuation of RNA viruses by redirecting their evolution in sequence space. *Nat Microbiol*. 2017; 2: nmicrobiol201788. <https://doi.org/10.1038/nmicrobiol.2017.88> PMID: 28581455
32. Lauring AS, Acevedo A, Cooper SB, Andino R. Codon usage determines the mutational robustness, evolutionary capacity and virulence of an RNA virus. *Cell Host Microbe*. 2012; 12: 623–632. <https://doi.org/10.1016/j.chom.2012.10.008> PMID: 23159052