medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Identification of a molnupiravir-associated mutational signature in SARS-CoV-2 sequencing databases

Theo Sanderson [1,✉], Ryan Hisner [2], I'ah Donovan-Banfield [3,4], Thomas Peacock [5], and Christopher Ruis [6,7,8,✉]

[1]Francis Crick Institute, London, UK; [2]Independent researcher, USA; [3]Department of Infection Biology and Microbiomes, Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK; [4]NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Liverpool, UK; [5]Department of Infectious Disease, Imperial College London, London, UK; [6]Molecular Immunity Unit, University of Cambridge Department of Medicine, MRC-Laboratory of Molecular Biology, Cambridge, UK; [7]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK; [8]Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge, UK

Molnupiravir, an antiviral medication that has been widely used against SARS-CoV-2, acts by inducing mutations in the virus genome during replication. Most random mutations are likely to be deleterious to the virus, and many will be lethal. Molnupiravir-induced elevated mutation rates have been shown to decrease viral load in animal models. However, it is possible that some patients treated with molnupiravir might not fully clear SARS-CoV-2 infections, with the potential for onward transmission of molnupiravir-mutated viruses. We set out to systematically investigate global sequencing databases for a signature of molnupiravir mutagenesis. We find that a specific class of long phylogenetic branches appear almost exclusively in sequences from 2022, after the introduction of molnupiravir treatment, and in countries and age-groups with widespread usage of the drug. We calculate a mutational spectrum from the AGILE placebo-controlled clinical trial of molnupiravir and show that its signature, with elevated G-to-A and C-to-T rates, largely corresponds to the mutational spectrum seen in these long branches. Our data suggest a signature of molnupiravir mutagenesis can be seen in global sequencing databases, in some cases with onwards transmission.

Correspondence: ***theo.sanderson@crick.ac.uk cr628@cam.ac.uk***

Introduction

Molnupiravir is an antiviral drug, licensed in some countries for the treatment of COVID-19. In the body, molnupiravir is ultimately converted into a nucleotide-analog, molnupiravir triphosphate (MTP)[1]. MTP is capable of being incorporated into RNA during strand synthesis, particularly by viral RNA-dependent RNA polymerases, where it can result in errors of sequence fidelity during viral genome replication. These errors in RNA replication result in many viral progeny that are non-viable, and so reduce the virus's effective rate of growth – molnupiravir was shown to reduce viral replication in 24 hours by 880-fold in vitro, and to reduce viral load in animal models (Rosenke et al., 2021). Molnupiravir initially showed some limited efficacy as a treatment for COVID-19 (Jayk Bernal et al., 2022; Extance, 2022), but subsequent larger clinical trials found that molnupiravir did not reduce hospitalisation or death rates in high risk groups (Butler, 2022). As one of the first orally bioavailable antivirals on the market, molnupiravir has been widely adopted by many countries, most recently China (Reuters, 2022). However, recent trial results and the approval of more efficacious antivirals have since led to several countries recommending against molnupiravir usage on the basis of limited effectiveness (NICE Guidance ; NC19CET, 2022).

MTP appears to be incorporated into nascent RNA primarily by acting as an analogue of cytosine (C), pairing opposite guanine (G) bases (Fig. 1). However, once incorporated, the molnupiravir (M)-base can transition into an alternative tautomeric form which resembles uracil (U) instead. This means that in the next round of replication, to give the positive-sense SARS-CoV-2 genome, the M base can pair with adenine (A), resulting in a G-to-A mutation, as shown in Fig. 2. Incorporation of MTP can also occur during the second step synthesis of the positive-sense genome. In this case, an initial positive-sense C correctly pairs with a G in the first round of replication, but this G then pairs with an M base during positive-sense synthesis. In the next round of replication this M can then pair with A, which will re-

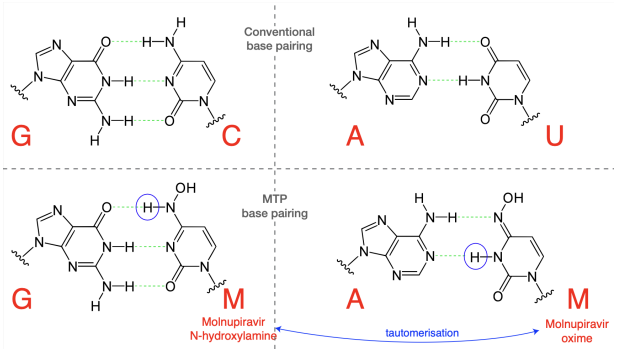

Figure 1. Molnupiravir triphosphate can assume multiple tautomeric forms. The N-hydroxylamine form resembles cytosine, while the oxime form more closely resembles uracil. They therefore pair with guanine and adenine respectively. (Figure adapted in part from Malone and Campbell (2021).)

[1]MTP is also known as β-D-N[4]-hydroxycytidine triphosphate (NHC-TP).

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

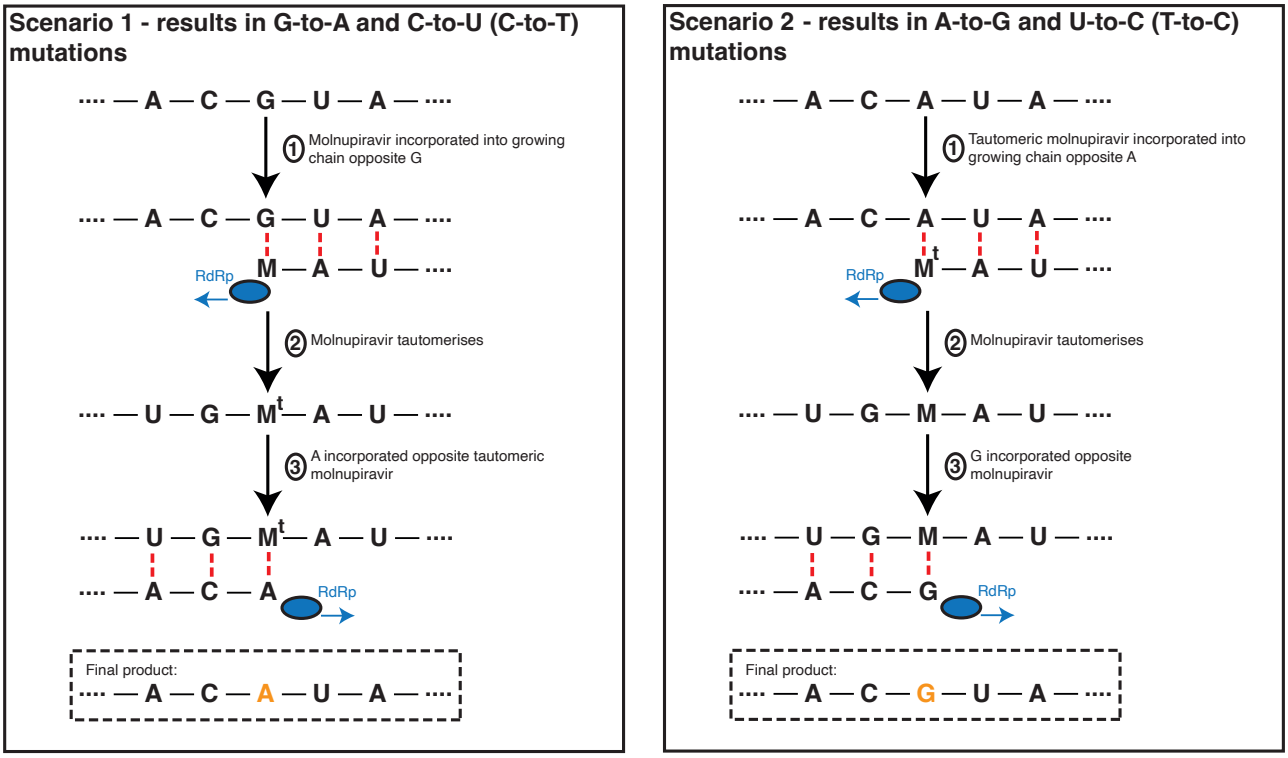

Figure 2. Molnupiravir drives G-to-A and C-to-U (C-to-T) mutations, and to a lesser extent A-to-G and T-to-U (T-to-C) mutations
In the most common scenario, shown on the left-hand side, M is incorporated opposite G nucleotides. It can then pair to A in subsequent replication, creating a G-to-A mutation. If the original G resulted from negative strand synthesis from a coding-strand C, then the G-to-A change ultimately creates a C-to-U coding change (Fig. S1). In the second, less common, scenario shown at the right hand side M is initially incorporated by pairing with A, which can result in an A-to-G mutation or, if the original A came from a coding U, a U-to-C mutation.

sult in a U in the final positive sense genome, with the overall process producing a C-to-U mutation (Fig. S1).

The free nucleotide MTP is less prone to tautomerisation than when incorporated into RNA, and so this directionality of mutations is the most likely (Gordon et al., 2021). However it is also possible for some MTP to bind, in place of U, to A bases and undergo the above processes in reverse, causing A-to-G and U-to-C mutations (Scenario 2 in Fig. 2, also Fig. S1).

It has been proposed that many major SARS-CoV-2 variants emerged from long-term chronic infections. This model explains several peculiarities of variants such as a general lack of genetic intermediates, rooting with much older sequences, long phylogenetic branch lengths, and the level of convergent evolution with known chronic infections (Rambaut et al., 2020; Viana et al., 2022; Hill et al., 2022; Harari et al., 2022).

During analysis of long phylogenetic branches in the SARS-CoV-2 tree, recent branches exhibiting potential hallmarks of molnupiravir-driven mutagenesis have been noted, including clusters of sequences indicating onward transmission. We therefore aimed to systematically identify sequences that might have been influenced by molnupiravir and characterise their mutational profile to examine the extent to which these signatures appear in global sequencing databases.

Results

During examination of SARS-CoV-2 phylogenetic branches with large numbers of mutations, a subset that contained skewed ratios of mutation classes and very few transversion substitutions were identified (Hisner, 2022). These patterns appeared very different to typical SARS-CoV-2 mutational spectra (Ruis et al., 2022; Bloom et al., 2022; Masone et al., 2022; Tonkin-Hill et al., 2021).

To investigate this pattern more systematically we analysed a mutation-annotated tree, derived from McBroome et al. (2021), containing >13 million SARS-CoV-2 sequences from GISAID (Elbe and Buckland-Merrett, 2017) and the INSDC databases (Cochrane et al., 2011). For each branch of the tree we counted the number of each substitution class (A-to-T, A-to-G, etc. – we use T instead of U for the remainder of this manuscript as in sequencing databases). Filtering this tree to branches involving at least 20 substitutions, and plotting the proportion of substitution types revealed a region of this space with higher G-to-A and almost exclusively transition substitutions[2], that only contained branches sampled in 2022 (Fig. 3A), suggesting some change (either biological or technical) had resulted in a

[2] "Transition" substitutions are: G-to-A, A-to-G, C-to-T, T-to-C. Other substitutions are known as "transversions".

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

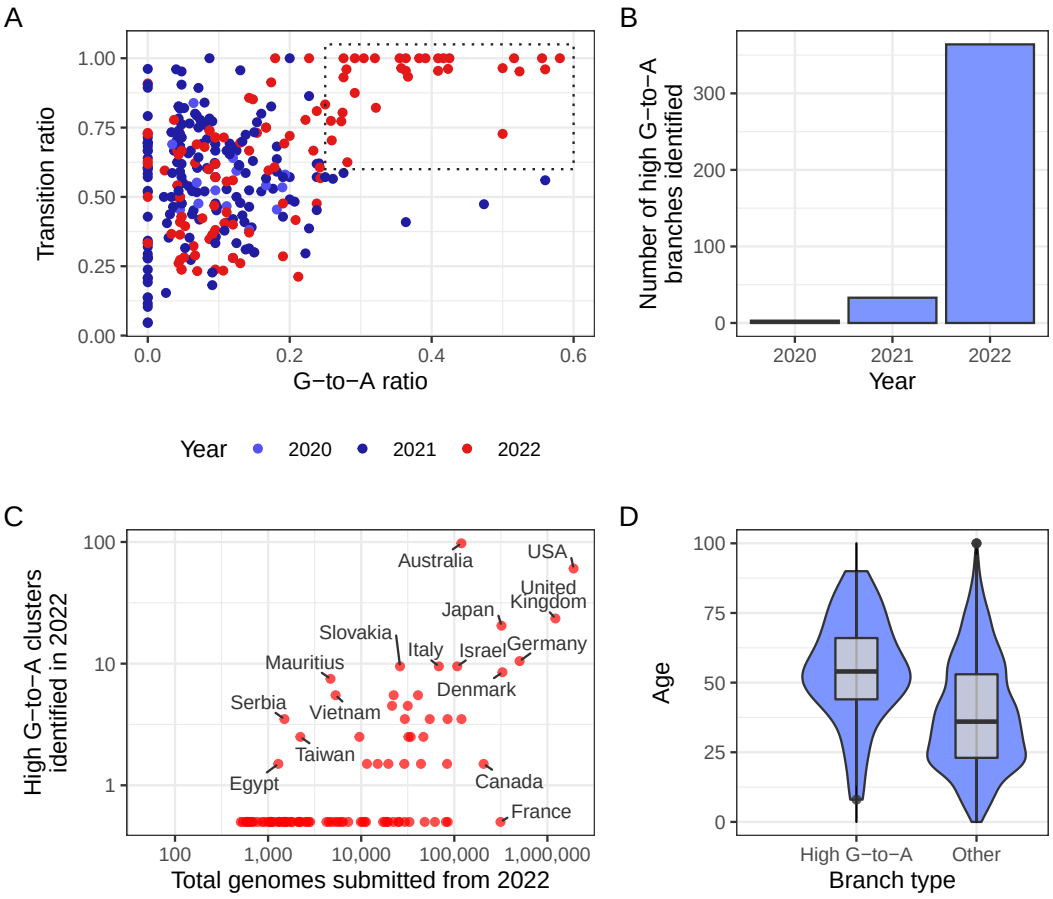

Figure 3. A new mutational signature emerged with high G-to-A and high transition ratio emerged in 2022 in some countries in global sequencing databases (**A**) Each point in this scatterplot represents a branch from the mutation-annotated tree with >20 substitutions. Points are positioned according the proportion of the branch's mutations that are G-to-A (x) or any transition mutation (y), and coloured by the year in which they occurred. A boxed region with higher G-to-A and almost exclusively transition mutations occurs only in 2022. (**B**) A count of the number of branches which satisfy a specific criterion (G-to-A ratio >= 25%, C-to-T ratio >= 20% transition ratio > 95%, total mutations > 10). (**C**) A comparison of number of total genomes with number of identified high G-to-A clusters (using the same criterion as B). Note logarithmic axes (semi-log, with the truncated line representing zero). For example, Australia has 97 clusters from a total of 119,194 genomes, whereas France has 0 clusters from 313,680 genomes. (**D**) A comparison of the age distributions for clusters with >10 mutations from the USA, partitioned by whether or not they satisfy the high G-to-A criterion. High G-to-A clusters correspond to older individuals.

new mutational signature.

Noticing that this signature also involved a high proportion of C-to-T mutations, we created a criterion for branches of interest, which we refer to as "high G-to-A" branches: we selected branches involving at least 10 substitutions, of which at least 25% were G-to-A, at least 20% were C-to-T and at most 5% were transversions. Again, these branches were almost all sampled in 2022 (Fig. 3B). The branches were predominantly sampled from a small number of countries, which could not be explained by differences in sequencing efforts (Fig. 3C, Table 1). Many countries which exhibited a high proportion of high G-to-A branches use molnupiravir: >380,000 prescriptions had occurred in Australia by the end of 2022 (Department of Aged Care Webinar, 2022), >30,000 in the UK (NHS, 2023; Butler, 2022), and >240,000 in the US within the early months of 2022 (Gold et al., 2022). Countries with high levels of total sequencing but a low number of G-to-A branches (Canada, France) have not authorised the use of molnupiravir (Goverment of Canada, 2022; Spencer et al., 2021). Age metadata from the US showed a significant bias towards patients with older ages for these high G-to-A branches, compared to control branches with similar numbers of mutations but without selection on substitution-type (Fig. 3D). Where age data was available in Australia it also identified long branches primarily in an aged population. This is consistent with the prioritised use of molnupiravir to treat older individuals, who are at greater risk from severe infection, in these coun-

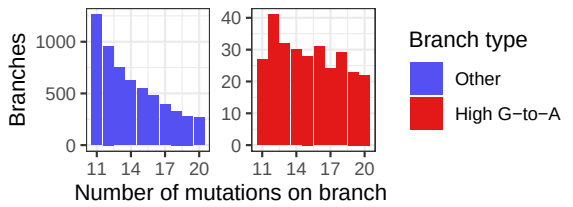

Figure 4. Branch length distribution for high G-to-A branches compared to other branches
Plot is limited to branch lengths greater than 10, in order to be confident in the branch type, and to branch lengths less than 20 due to the exclusion of long branch samples from the UShER MAT. Branches are filtered to those from 2022.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

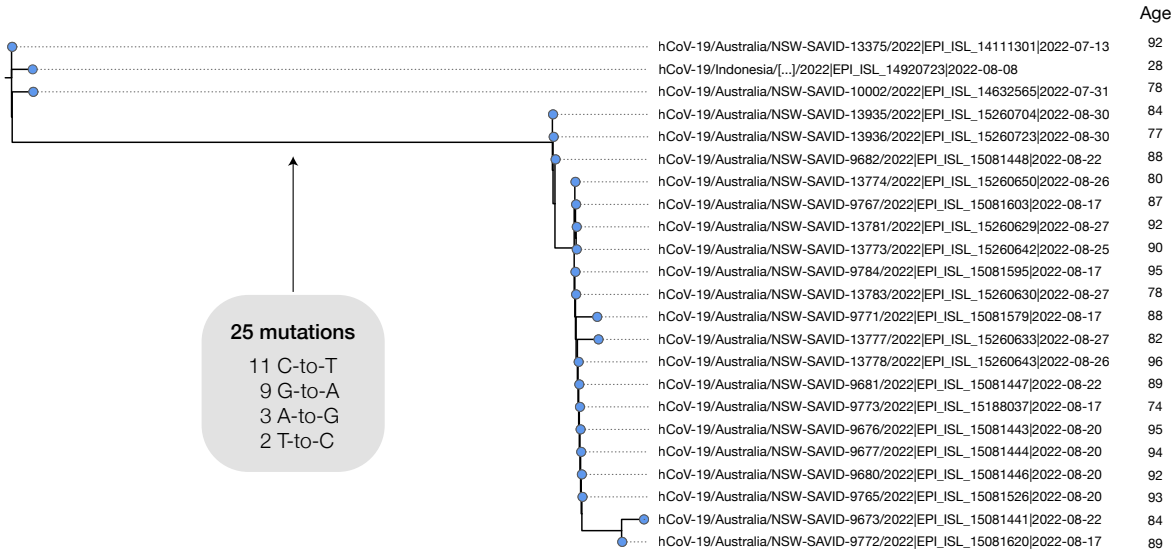

Figure 5. A cluster of 20 individuals emerging from a high G-to-A mutation event
This cluster involves a saltation of 25 mutations occuring within a period of perhaps under a month, all of which are transition substitutions, with an elevated G-to-A rate. Sequences are annotated with age metadata suggestive of an outbreak in an aged-care facility.

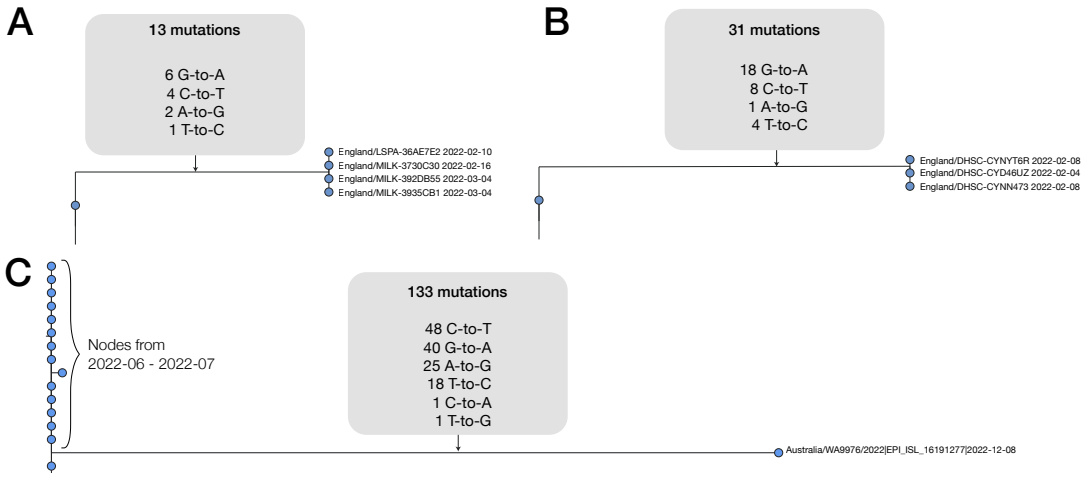

Figure 6. Phylogenetic trees for three high G-to-A events
(**A**) A cluster of four sequences from the UK from Feb-March 2022 with 13 shared muations with the high G-to-A signature. (**B**) A cluster of four sequences from the UK from Feb 2022 with 31 shared muations with the high G-to-A signature. (**C**) A singleton sequence from Australia with a high G-to-A signature and a total of 133 mutations. Just 2 of the 133 mutations observed are transversions and transitions include numerous G-to-A events.

tries. In Australia, molnupiravir was pre-placed in aged-care facilities, and it was recommended that it be considered for all patients aged 70 or older, with or without symptoms (Australian Department of Health and Aged Care, 2022).

We found that high G-to-A branches had a different distribution of branch lengths from other types of branches, with an enrichment for longer branch lengths (Fig. 4). We next sought to see whether these high G-to-A branches were associated with a different mutation rate to other branches of similar length. We assigned a date to each node of the tree using Chronumental (Sanderson, 2021) and observed that the branch length measured in time was shorter for branches with a high G-to-A signature than for branches of matched length without this signature (Fig. S2), suggesting an increased mutation rate.

Most of the long branches sampled have just a single descendant tip sequence in sequencing databases, but in some cases branches have given rise to clusters with a significant number of descendant sequences. For example, a cluster in Australia in August 2022 involves 20 tip sequences, with distinct age metadata indicating that they truly derive from multiple individuals (Fig. 5). This cluster involves 25 substitutions in the main branch of which all are transitions, 44% are C-to-T and 36% are G-to-A. Closely related outgroups date from July 2022, suggesting that these mutations emerged in a period of 1-2 months. There are many other examples of high G-to-A branches with multiple descendant sequences, including from the United Kingdom (Fig. 6A,B).

During the construction of the daily-updated mutation-

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

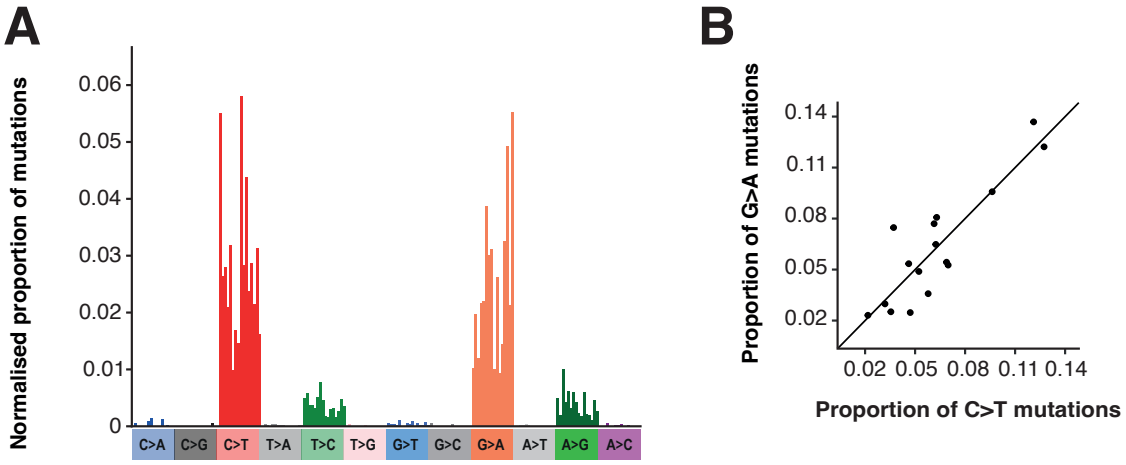

Figure 7. Mutational spectrum from long phylogenetic branches in global sequencing databases
(**A**) Single base substitution (SBS) spectrum of long phylogenetic branches. (**B**) Comparison of contextual mutation preferences in C-to-T and G-to-A mutations on long phylogenetic branches. The proportion of C-to-T or G-to-A mutations in the equivalent (reverse) contexts are plotted. There is a significant correlation, showing similar contextual preferences for C-to-T and G-to-A mutations. Proportions are normalised for the number of times the context occurs in the genome.

annotated tree (McBroome et al., 2021), samples highly divergent from the existing tree are excluded. This is a necessary step given the technical errors in some SARS-CoV-2 sequencing data, but it also means that highly divergent molnupiravir-induced sequences might be excluded. To examine this effect, we created a comprehensive mutation-annotated-tree for Australia, allowing identification of even the most divergent sequences in this subset. This analysis allowed the identification of mutational events involving up to 130 substitutions (Fig. 6C, Fig. S3), with the same signature of elevated G-to-A mutation rates and almost exclusively transition substitutions. The cases we identified with these very high numbers of mutations involved single sequences, and could represent sequences resulting from chronically infected individuals who have been treated with multiple courses of molnupiravir.

We next assessed the mutational spectrum (the patterns of contextual nucleotide substitutions) on these branches. The spectrum we identified (Fig. 7A) is dominated by G-to-A and C-to-T transition mutations with smaller contributions from A-to-G and T-to-C transitions. This pattern is consistent with the known mechanisms of action for molnupiravir (Fig. 2). These transitions exhibit preference for particular surrounding nucleotide contexts, for example G-to-A mutations occur most commonly in TGT and TGC contexts. This may represent a preference for molnupiravir binding adjacent to particular surrounding nucleotides, a preference of the viral RdRp to incorporate molnupiravir adjacent to specific nucleotides, or a preference of the viral proofreading exonuclease to remove molnupiravir in specific contextual surroundings.

The dominance of both C-to-T and G-to-A mutations in the spectrum is likely due to molnupiravir-induced G-to-A mutations during synthesis of different strands during virus replication. Incorporation of molnupiravir during negative strand synthesis will result in G-to-A mutations in the virus consensus sequence while incorporation during positive strand synthesis will be read as C-to-T mutations (Fig. S1). Consistent with this, we observe a strong positive correlation between the mutational biases in equivalent contextual patterns within C-to-T and G-to-A mutations (Pearson's $r = 0.88$, 95% CI 0.68-0.96, $p < 0.001$, Fig. 7B) (e.g. a C-to-T mutation in the ACG context on one strand is the equivalent of a G-to-A mutation in the CGT context on the other strand).

To compare the observed signatures to mutations in known molnupiravir-exposed individuals, we re-examined a genomic dataset from the AGILE Phase IIa clinical trial (Donovan-Banfield et al. (2022), NCT04746183). For the first time we analysed the spectrum of likely molnupiravir-induced mutations by comparing day 1 samples (taken just before treatment initiation) to day 5 samples from the same patient. This dataset had the advantage that it also included individuals treated with placebo, providing a control for mutational spectra in the absence of molnupiravir. We found that molnupiravir-treated patients exhibited significantly higher mutational burdens than patients treated with placebo (ANOVA $p < 0.001$, Fig. S6). The spectrum of mutations was highly different between placebo and molnupiravir (Fig. 8A, cosine similarity between spectra = 0.68). Assuming that the mutational processes within the placebo patients also occurred in the molnupiravir treatment, we subtracted the placebo spectrum to obtain the mutations specifically induced by molnupiravir (Fig. 8A). Again, we found a significant enrichment of transition mutations (Fig. 8B-C).

We used the calculated spectrum to examine whether the long phylogenetic branches identified above are likely to be molnupiravir-driven. The overall patterns of mutations in the long phylogenetic branches are qualitatively similar to the AGILE trial patients treated with molnupiravir, with a preponderance of transition mutations, most commonly C-to-T and G-to-A (Fig. 7, Fig. 8).

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

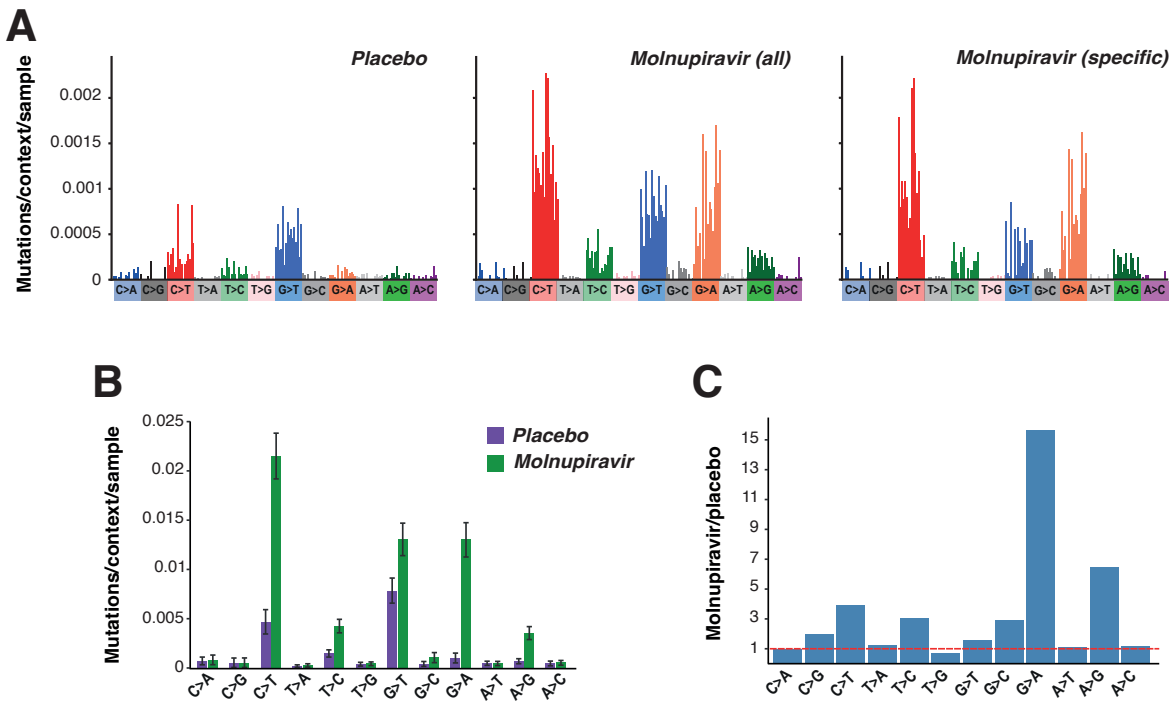

Figure 8. Molnupiravir mutational spectra from the AGILE clinical trial
(A) SBS spectra of mutations in patients treated with placebo or molnupiravir in the AGILE trial. The right hand panel shows the mutations induced by molnupiravir, calculated by subtracting the placebo spectrum (left panel) from the spectrum of all mutations in molnupiravir treated patients (centre panel). Spectra are shown as the number of mutations per available context per patient. **(B)** Comparison of the number of each mutation type (summed across all contexts) between placebo and molnupiravir treatments. Error bars show confidence intervals from mutation bootstrapping (see methods). **(C)** The ratio of each mutation type between treatments was calculated by dividing the mutational burden in molnupiravir by that in placebo. The red dashed line shows an equal burden.

The contextual patterns within each transition mutation are highly similar between the known molnupiravir spectrum and the long phylogenetic branches (Fig. S7). This suggests a shared driver of transition mutations and therefore supports the long phylogenetic branches being driven by treatment with molnupiravir.
While the transition patterns are highly similar, the AGILE trial molnupiravir spectrum contains a high rate of G-to-T mutations that is not present in the long phylogenetic branches (Fig. 7, Fig. 8). This high rate is also present in the placebo-treated patients, although the rate appears higher in molnupiravir treatment (Fig. 8A). The rate of G-to-T mutations in both placebo and molnupiravir groups appears higher than that calculated from within patient mutations from untreated individuals (Tonkin-Hill et al. (2021), Fig. S8), for reasons that are not clear.

Discussion

We have shown a variety of lines of evidence that together suggest that a signature of molnupiravir treatment is visible in global sequencing databases. We identified a set of long phylogenetic branches that exhibit a high number of transition mutations. The number of these branches increased dramatically in 2022 and they are enriched for countries and age groups known to be exposed to molnupiravir. Mutation rates on these branches, were elevated, consistent with a recent study from Fountain-Jones et al. (2022) in immunocompromised patients. The branches exhibited a mutational spectrum that is highly similar to that in patients known to be treated with molnupiravir. The sequencing data suggest that in at least some cases, viruses with a large number of molnupiravir-induced substitutions have been transmitted to other individuals, at least in a limited manner.

Molnupiravir's mode of action is often described using the term "error catastrophe" – the concept that there is an upper limit on the mutation rate of a virus beyond which it is unable to maintain self-identity (Eigen, 1971). This model has been criticised on its own terms (Summers and Litwin, 2006), but is particularly problematic in the case of molnupiravir treatment. The model assumes a steady-state condition, with the mutation rate fixed at a particular level. The threshold for error catastrophe is the mutation rate at which, according to the model, the starting sequence will ultimately be lost after an infinite time at this steady-state. However molnupiravir treatment does not involve a steady state, but a temporarily elevated mutation rate over a short treatment period. Therefore there isn't any particular threshold point, or well-defined "catastrophe" condition. We suggest that the use of this term in the context of mutagenic antiviral drugs is unhelpful. It is enough to think of these drugs as acting through mutagenesis to reduce the number of viable progeny that each virion is able to produce – particularly given that much of the reduction in fitness

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

will be due to "single-hit" lethal events (Summers and Litwin, 2006).

New variants of SARS-CoV-2 are generated through acquisition of mutations that enhance properties including immune evasion and intrinsic transmissibility (Telenti et al., 2022; Carabelli et al., 2023). The impact of molnupiravir treatment on the trajectory of variant generation and transmission is difficult to predict. On the one hand, molnupiravir increases the amount of sequence diversity in the surviving viral population in the host and this might be expected to provide more material for selection to act on during intra-host evolution towards these properties that increase fitness. However, a high proportion of induced mutations are likely to be deleterious or neutral, and it is necessary to consider the counterfactual to molnupiravir treatment. As molnupiravir results in a modest reduction in viral load in treated patients (Khoo et al., 2022), it is possible that in the absence of treatment the total viral load would be higher and chronic infections might persist for longer. Variants generated through chronic infections might be fitter than those that have accumulated mutations during molnupiravir treatment, albeit taking a much longer period of time to accumulate the same number of mutations and therefore usually being derived from older, rather than contemporary lineages. At the time of writing we have not identified a molnupiravir-implicated cluster that had spread to more than 21 individuals.

There are some limitations of our work. Detecting a particular branch as involving a molnupiravir-like signature is a probabilistic rather than absolute judgement: where molnupiravir creates just a handful of mutations (which trial data suggests is often the case), branch lengths will be too small to assign the cause of the mutations with confidence. We therefore limited our analyses here to long branches. This approach may also fail to detect branches which feature a substantial number of molnupiravir-induced mutations alongside a considerable number of mutations from other causes (which might occur in chronic infections). We discovered drastically different rates of molnupiravir-associated sequences by country, and suggest that this reflects in part whether, and how, molnupiravir is used in different geographical regions – however there will also be contributions from the rate at which genomes are sequenced in settings where molnupiravir is used. For example, if molnupiravir is used primarily in aged-care facilities and viruses in these facilities are significantly more likely to be sequenced than those in the general community this will elevate the ascertainment rate of such sequences. Furthermore, it is likely that some included sequences were specifically analysed due to representing continued test positivity after molnupiravir treatment as part of specific studies. Such effects are likely to differ based on sequencing priorities in different locations.

We would recommend public health authorities in countries showing these patterns perform investigations to determine if these sequences or clusters can indeed be directly linked back to use of molnupiravir. These data will be useful for ongoing assessments of the risks and benefits of this treatment, and may guide the future development of mutagenic agents as antivirals, particularly for viruses with high mutational tolerances such as coronaviruses.

Methods

Processing of mutation-annotated tree

To identify clusters in global sequence databases with a molnupiravir associated signature we analysed a regularly updated mutation-annotated tree built using UShER (Turakhia et al., 2021) with almost all global data – a version of the McBroome et al. (2021) tree. We parsed the tree using a custom script adapted from Taxonium Tools (Sanderson, 2022). The script added metadata from sequencing databases to each node, then passed these metadata to parent nodes using simple heuristics: (1) a parent node was annotated with a year if all of its descendants were annotated with that year, (2) a parent node was annotated with a particular country if all of its descendants were annotated with that country, (3) a parent node was annotated with the mean age of its (age-annotated) descendants.

Mutation rate analysis

We used Chronumental (Sanderson, 2021) to assign dates to each node in the mutation-annotated tree. We ran Chronumental for 300 steps, then extracted length of time in days from the output Newick tree, and compared against the number of mutations on the branch, splitting by whether the branch satisfied our criteria to be a high G-to-A branch.

Generation of cluster trees

For the cluster of 20 individuals, we observed small imperfections in UShER's representation of the mutation-annoated tree within the cluster resulting from missing coverage at some positions. We therefore recalculated the tree that we display here. We took the 20 sequences in the cluster, and the three closest outgroup sequences, we aligned using Nextclade (Aksamentov et al., 2021), calculated a tree using iqtree (Minh et al., 2020) and reconstructed the mutation-annotated tree using TreeTime (Sagulenko et al., 2018). We visualised the tree using FigTree (Rambaut, 2018). We also visualised Fig. S3 with NextStrain (Hadfield et al., 2018).

Adding back excluded divergent sequences to the mutation-annotated tree

We used GISAIDR (Wirth and Duchene, 2022) to download all Australian sequences in 2022 from the GISAID database. We then filtered those based on sequence ID to those absent from the existing mutation-anotated tree (MAT). We aligned these sequences to the Hu-1 reference using flowalign. We pruned the existing MAT to

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Table 1. Number of high G-to-A (G-to-A ratio >= 25%, C-to-T ratio >= 20% transition ratio > 95%, total mutations > 10) identified by country in 2022, set against total number of genomes. Only countries with >10,000 genomes are included.

Country	High G-to-A branches in 2022	Total genomes in 2022
Australia	97	119,194
USA	60	1,911,997
United Kingdom	23	1,218,724
Japan	20	321,520
Germany	10	503,014
Israel	9	107,477
Italy	9	68,229
Slovakia	9	26,124
Denmark	8	329,165
New Zealand	5	22,345
Russia	5	40,783
Czech Republic	4	31,610
Thailand	4	21,459
India	3	119,578
Luxembourg	3	29,435
South Korea	3	54,505
Spain	3	84,749
Austria	2	34,028
Indonesia	2	32,141
Ireland	2	46,603
Belgium	1	83,648
Canada	1	206,718
Philippines	1	11,513
Poland	1	43,865
Slovenia	1	29,062
South Africa	1	15,043
Turkey	1	19,592
Brazil	0	85,059
Chile	0	19,336
Colombia	0	11,257
Croatia	0	22,160
Finland	0	17,978
France	0	313,680
Greece	0	10,339
Hong Kong	0	10,901
Malaysia	0	25,605
Mexico	0	33,008
Netherlands	0	61,905
Norway	0	29,381
Peru	0	25,000
Portugal	0	19,135
Singapore	0	17,169
Sweden	0	81,326
Switzerland	0	48,087

retain only Australian sequences, and then added all sequences missing from the tree using UShER (Turakhia et al., 2021), without filtering on parsimonious placements or path length, to achieve a complete MAT for Australia.

Calculation of mutational spectrum of long phylogenetic branches

To calculate the SBS mutational spectrum of the long phylogenetic branches, we extracted the path of mutations leading to each branch from the 2022-12-18 UShER phylogenetic tree. Internal and tip branches in these paths containing at least ten mutations were identified and their mutations extracted. The context of each mutation was identified using the Wuhan-Hu-1 genome (accession NC_045512.2), incorporating mutations acquired earlier in the path. Mutation counts were rescaled by genomic content by dividing the number of mutations by the count of the starting triplet in the Wuhan-Hu-1 genome. MutTui (`https://github.com/chrisruis/MutTui`) was used to rescale and plot mutational spectra.

Calculation of molnupiravir and placebo mutational spectra using AGILE trial data

We calculated molnupiravir and placebo SBS spectra using previously published variant data (Donovan-Banfield et al., 2022). We used deep sequencing data from samples collected on day one (pre-treatment) and day five (post-treatment) from 65 patients treated with placebo and 58 patients treated with molnupiravir. For each patient, we used the consensus sequence of the day one sample as the reference sequence and identified mutations as variants in the day five sample away from the patient reference sequence in at least 5% of reads at genome sites with at least 100-fold coverage. The surrounding nucleotide context of each mutation was identified from the patient reference sequence.

We converted mutation counts to mutational burdens by dividing each mutation count by the number of the starting triplet in the Wuhan-Hu-1 genome (accession NC_045512.2) and by the number of patients in the treatment group (65 for placebo and 58 for molnupiravir). This therefore rescales by the number of opportunities for each mutation to occur across the genome and the number of patients in each group.

To ensure that any spectrum differences between placebo and molnupiravir treatments are not due to previously observed differences in spectrum between SARS-CoV-2 variants (Ruis et al., 2022; Bloom et al., 2022), we compared the distribution of variants between the treatments (Fig. S4). The distributions were highly similar.

To calculate the total mutational burden of each mutational class, we summed the mutational burden of the 16 contextual mutations within the class. Confidence intervals were calculated by bootstrapping mutations. Here, the original set of mutations within the treatment was resampled with replacement before rescaling by triplet availability and number of patients. 1000 bootstraps were run and 95% confidence intervals calculated.

To enable comparison with an additional SARS-CoV-2 dataset containing mutations acquired during within host infections, we obtained variants from a previous deep sequencing study Tonkin-Hill et al. (2021). We identified the surrounding nucleotide context of each mutation using the Wuhan-Hu-1 genome and rescaled by context availability.

Comparison of mutational spectra

The cosine similarity between placebo and molnupiravir spectra was calculated using MutTui

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

(https://github.com/chrisruis/MutTui).
We compared contextual patterns within each transition mutation between the long phylogenetic branch and AGILE molnupiravir spectra through regression of the proportion of mutations within the mutational class that are in each context. To assess significance of the correlation, we randomised the proportions within the mutational class within each spectrum and recalculated the correlation. 1000 randomisations were carried out and the p-value calculated as the proportion of randomisations with a correlation at least as large as in the real data.

ACKNOWLEDGEMENTS

We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. We are also very grateful to everyone who has contributed to the generation of the genomes that have been deposited in the INSDC databases, on which this research is also based. We thank Angie Hinrichs and colleagues for access to an UShER mutation-annotated tree built with all available genomic data. We thank Jesse Bloom, Michael Lin, Richard Neher, and Kelley Harris for useful discussions. This preprint uses a LaTeX template from Stephen Royle and Ricardo Henriques.

DATA AVAILABILITY

No new primary data was generated for this study. We used data from international sequencing databases (Elbe and Buckland-Merrett, 2017; Cochrane et al., 2011), and from the AGILE clinical trial (Donovan-Banfield et al., 2022), where genomic data were obtained from BioProject PRJNA854613 at the SRA. Our GitHub repository is available at https://github.com/theosanderson/molnupiravir.
The findings of this study are based on metadata associated with 14,449,737 sequences available on GISAID up to December 2022, and accessible at 10.55876/gis8.230110wz and 10.55876/gis8.230110db (see also, Supplemental Tables). The findings of this study are also based on 6,594,478 sequences from INSDC – authors, metadata, and sequences are available here. Data present in both databases are deduplicated before the construction of the mutation-annotated tree.

AUTHOR CONTRIBUTIONS

RH identified initial branches, and their likely connection to molnupiravir. TS performed analyses of mutation-annotated tree and global metadata. CR performed all mutational spectra analyses. ID-B created bioinformatic pipelines for the AGILE trial data. All authors participated in mansuscript writing.

FUNDING

TS was supported by the Wellcome Trust (210918/Z/18/Z) and the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001043), the UK Medical Research Council (FC001043), and the Wellcome Trust (FC001043). This research was funded in whole, or in part, by the Wellcome Trust [210918/Z/18/Z, FC001043]. For the purpose of Open Access, the authors have applied a CC-BY public copyright licence to any Author Accepted Manuscript resulting from this preprint.
ID-B is supported by PhD funding from the National Institute for Health and Care Research (NIHR) Health Protection Research Unit (HPRU) in Emerging and Zoonotic Infections at University of Liverpool in partnership with Public Health England (PHE) (now UKHSA), in collaboration with Liverpool School of Tropical Medicine and the University of Oxford (award 200907). The views expressed are those of the authors and not necessarily those of the Department of Health and Social Care or NIHR. Neither the funders or trial sponsor were involved in the study design, data collection, analysis, interpretation, nor the preparation of the manuscript.
TP was funded by the G2P-UK National Virology Consortium funded by the MRC (MR/W005611/1).
CR was supported by a Fondation Botnar Research Award (Programme grant 6063) and UK Cystic Fibrosis Trust (Innovation Hub Award 001).

Bibliography

Aksamentov, I., Roemer, C., Hodcroft, E., and Neher, R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. *Journal of Open Source Software*, 6(67):3773, nov 30 2021.

Australian Department of Health and Aged Care, 2022. Use of Lagevrio (molnupiravir) in residential aged care. URL https://www.health.gov.au/sites/default/files/documents/2022/07/coronavirus-covid-19-use-of-lagevrio-molnupiravir-in-residential-aged-care.pdf. Accessed: 2022-01-01.

Bloom, J. D., Beichman, A. C., Neher, R. A., and Harris, K. Evolution of the SARS-CoV-2 mutational spectrum. *bioRxiv*, Nov. 2022.

Butler, J. Molnupiravir plus usual care versus usual care alone as early treatment for adults with COVID-19 at increased risk of adverse outcomes (PANORAMIC): an open-label, platform-adaptive randomised controlled trial. *Lancet*, Dec. 2022.

Carabelli, A. M., Peacock, T. P., Thorne, L. G., Harvey, W. T., Hughes, J., COVID-19 Genomics UK Consortium, Peacock, S. J., Barclay, W. S., de Silva, T. I., Towers, G. J., and Robertson, D. L. Sars-CoV-2 variant biology: immune escape, transmission and fitness. *Nature Reviews Microbiology*, 2023.

Cochrane, G., Karsch-Mizrachi, I., Nakamura, Y., and on behalf of the International Nucleotide Sequence Database Collaboration. The international nucleotide sequence database collaboration, 2011.

Department of Aged Care Webinar, 2022. https://www.health.gov.au/resources/webinars/covid-19-response-update-for-primary-care-15-december-2022?language=en, 2022.

Donovan-Banfield, I., Penrice-Randal, R., Goldswain, H., Rzeszutek, A. M., Pilgrim, J., Bullock, K., Saunders, G., Northey, J., Dong, X., Ryan, Y., Reynolds, H., Tetlow, M., Walker, L. E., FitzGerald, R., Hale, C., Lyon, R., Woods, C., Ahmad, S., Hadjiyiannakis, D., Periselneris, J., Knox, E., Middleton, C., Lavelle-Langham, L., Shaw, V., Greenhalf, W., Edwards, T., Lalloo, D. G., Edwards, C. J., Darby, A. C., Carroll, M. W., Griffiths, G., Khoo, S. H., Hiscox, J. A., and Fletcher, T. Characterisation of SARS-CoV-2 genomic variation in response to molnupiravir treatment in the AGILE Phase IIa clinical trial. *Nat Commun*, 13(1):7284, Nov 2022.

Eigen, M. Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften*, 58(10):465–523, Oct. 1971.

Elbe, S. and Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall*, 1(1):33–46, Jan. 2017.

Extance, A. Covid-19: What is the evidence for the antiviral molnupiravir? *BMJ*, 377:o926, Apr. 2022.

Fountain-Jones, N. M., Vanhaeften, R., Williamson, J., Maskell, J., Chua, I.-L. J., Charleston, M., and Cooley, L. Antiviral treatments lead to the rapid accrual of hundreds of SARS-CoV-2 mutations in immunocompromised patients. dec 22 2022.

Gold, J. A. W., Kelleher, J., Magid, J., Jackson, B. R., Pennini, M. E., Kushner, D., Weston, E. J., Rasulnia, B., Kuwabara, S., Bennett, K., Mahon, B. E., Patel, A., and Auerbach, J. Dispensing of Oral Antiviral Drugs for Treatment of COVID-19 by Zip Code-Level Social Vulnerability - United States, December 23, 2021-May 21, 2022. *MMWR Morb Mortal Wkly Rep*, 71(25):825–829, Jun 2022.

Gordon, C. J., Tchesnokov, E. P., Schinazi, R. F., and Götte, M. Molnupiravir promotes SARS-CoV-2 mutagenesis via the RNA template. *J. Biol. Chem.*, 297(1), July 2021.

Goverment of Canada, 2022. Drug and vaccine authorizations for COVID-19: List of authorized drugs, vaccines and expanded indications. https://www.canada.ca/en/health-canada/services/drugs-health-products/covid19-industry/drugs-vaccines-treatments/authorization/list-drugs.html, 2022.

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., and Neher, R. A. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*, 34(23):4121–4123, may 22 2018.

Harari, S., Tahor, M., Rutsinsky, N., Meijer, S., Miller, D., Henig, O., Halutz, O., Levytskyi, K., Ben-Ami, R., Adler, A., Paran, Y., and Stern, A. Drivers of adaptive evolution during chronic SARS-CoV-2 infections. *Nat Med*, 28(7):1501–1508, Jul 2022.

Hill, V., Du Plessis, L., Peacock, T. P., Aggarwal, D., Colquhoun, R., Carabelli, A. M., Ellaby, N., Gallagher, E., Groves, N., Jackson, B., McCrone, J. T., O'Toole, A., Price, A., Sanderson, T., Scher, E., Southgate, J., Volz, E., Barclay, W. S., Barrett, J. C., Chand, M., Connor, T., Goodfellow, I., Gupta, R. K., Harrison, E. M., Loman, N., Myers, R., Robertson, D. L., Pybus, O. G., and Rambaut, A. The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evol*, 8(2):veac080, 2022.

Hisner, R. RE:Potential BA.2.3 sublineage with many mutations (singleton, Indonesia). https://github.com/cov-lineages/pango-designation/issues/1080#issuecomment-1250412876, 2022.

Jayk Bernal, A., Gomes da Silva, M. M., Musungaie, D. B., Kovalchuk, E., Gonzalez, A., Delos Reyes, V., Martín-Quirós, A., Caraco, Y., Williams-Diaz, A., Brown, M. L., Du, J., Pedley, A., Assaid, C., Strizki, J., Grobler, J. A., Shamsuddin, H. H., Tipping, R., Wan, H., Paschke, A., Butterton, J. R., Johnson, M. G., De Anda, C., and MOVe-OUT Study Group. Molnupiravir for oral treatment of covid-19 in nonhospitalized patients. *N. Engl. J. Med.*, 386(6):509–520, Feb. 2022.

Khoo, S. H., FitzGerald, R., Saunders, G., Middleton, C., Ahmad, S., Edwards, C. J., Hadjiyiannakis, D., Walker, L., Lyon, R., Shaw, V., Mozgunov, P., Periselneris, J., Woods, C., Bullock, K., Hale, C., Reynolds, H., Downs, N., Ewings, S., Buadi, A., Cameron, D., Edwards, T., Knox, E., Donovan-Banfield, I., Greenhalf, W., Chiong, J., Lavelle-Langham, L., Jacobs, M., Northey, J., Painter, W., Holman, W., Lalloo, D. G., Tetlow, M., Hiscox, J. A., Jaki, T., Fletcher, T., Griffiths, G., Paton, N., Hayden, F., Darbyshire, J., Lucas, A.,

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Lorch, U., Freedman, A., Knight, R., Julious, S., Byrne, R., Cubas Atienzar, A., Jones, J., Williams, C., Song, A., Dixon, J., Alexandersson, A., Hatchard, P., Tilt, E., Titman, A., Doce Carracedo, A., Chandran Gorner, V., Davies, A., Woodhouse, L., Carlucci, N., Okenyi, E., Bula, M., Dodd, K., Gibney, J., Dry, L., Rashid Gardner, Z., Sammour, A., Cole, C., Rowland, T., Tsakiroglu, M., Yip, V., Osanlou, R., Stewart, A., Parker, B., Turgut, T., Ahmed, A., Starkey, K., Subin, S., Stockdale, J., Herring, L., Baker, J., Oliver, A., Pacurar, M., Owens, D., Munro, A., Babbage, G., Faust, S., Harvey, M., Pratt, D., Nagra, D., and Vyas, A. Molnupiravir versus placebo in unvaccinated and vaccinated patients with early SARS-CoV-2 infection in the UK (AGILE CST-2): a randomised, placebo-controlled, double-blind, phase 2 trial. *Lancet Infect Dis*, Oct 2022.

Malone, B. and Campbell, E. A. Molnupiravir: coding for catastrophe. *Nat Struct Mol Biol*, 28 (9):706–708, Sep 2021.

Masone, D., Alvarez, M. S., and Polo, L. M. The SARS-CoV-2 mutation landscape is shaped before replication starts. Sept. 2022.

McBroome, J., Thornlow, B., Hinrichs, A. S., Kramer, A., De Maio, N., Goldman, N., Haussler, D., Corbett-Detig, R., and Turakhia, Y. A Daily-Updated database and tools for comprehensive SARS-CoV-2 Mutation-Annotated trees. *Mol. Biol. Evol.*, 38(12):5819–5824, Dec. 2021.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., and Lanfear, R. Iq-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, 37(5):1530–1534, feb 3 2020.

NC19CET. Taskforce updates molnupiravir guidance following PANORAMIC trial results. https://clinicalevidence.net.au/news/taskforce-updates-molnupiravir-guidance-following-panoramic-trial-results/, Dec. 2022. Accessed: 2023-1-6.

NHS, 2023. COVID Therapeutics Weekly Publication (week ending 1st January 2023). https://www.england.nhs.uk/statistics/statistical-work-areas/covid-therapeutics-antivirals-and-neutralising-monoclonal-antibodies/, 2022.

NICE Guidance . NICE recommends 3 treatments for COVID-19 in draft guidance. URL https://www.nice.org.uk/news/article/nice-recommends-3-treatments-for-covid-19-in-draft-guidance.

Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D. L., Volz, E., and Network, A. Preliminary genomic characterisation of an emergent sars-cov-2 lineage in the uk defined by a novel set of spike mutations, Dec 2020. URL https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563.

Rambaut, A. Figtree, 2018. URL http://tree.bio.ed.ac.uk/software/figtree/.

Reuters. China grants conditional approval for merck's COVID treatment. *Reuters*, Dec. 2022.

Rosenke, K., Hansen, F., Schwarz, B., Feldmann, F., Haddock, E., Rosenke, R., Barbian, K., Meade-White, K., Okumura, A., Leventhal, S., Hawman, D. W., Ricotta, E., Bosio, C. M., Martens, C., Saturday, G., Feldmann, H., and Jarvis, M. A. Orally delivered MK-4482 inhibits SARS-CoV-2 replication in the syrian hamster model. *Nat. Commun.*, 12, 2021.

Ruis, C., Peacock, T. P., Polo, L. M., Masone, D., Alvarez, M. S., Hinrichs, A. S., Turakhia, Y., Cheng, Y., McBroome, J., Corbett-Detig, R., Parkhill, J., and Andres Floto, R. Mutational spectra distinguish SARS-CoV-2 replication niches. Sept. 2022.

Sagulenko, P., Puller, V., and Neher, R. A. Treetime: Maximum-likelihood phylodynamic analysis. *Virus evolution*, 4(1):vex042, jan 8 2018. [Online; accessed 2023-01-13].

Sanderson, T. *bioRxiv*, oct 28 2021.

Sanderson, T. Taxonium, a web-based tool for exploring large phylogenetic trees. *Elife*, 11, Nov 2022.

Spencer et al., 2021. France cancels order for Merck's COVID-19 antiviral drug. https://www.reuters.com/world/europe/france-cancels-order-mercks-covid-19-antiviral-drug-2021-12-22/, 2021.

Summers, J. and Litwin, S. Examining the theory of error catastrophe. *J. Virol.*, 80(1):20–26, Jan. 2006.

Telenti, A., Hodcroft, E., and Robertson, D. The evolution and biology of SARS-CoV-2 variants. *Cold Spring Harbor Perspectives in Medicine*, 12(5):a041390, May 2022.

Tonkin-Hill, G., Martincorena, I., Amato, R., Lawson, A. R. J., Gerstung, M., Johnston, I., Jackson, D. K., Park, N., Lensing, S. V., Quail, M. A., Gonçalves, S., Ariani, C., Chapman, M. S., Hamilton, W. L., Meredith, L. W., Hall, G., Jahun, A. S., Chaudhry, Y., Hosmillo, M., Pinckert, M. L., Georgana, I., Yakovleva, A., Caller, L. G., Caddy, S. L., Feltwell, T., Khokhar, F. A., Houldcroft, C. J., Curran, M. D., Parmar, S., The COVID-19 Genomics UK (COG-UK) Consortium, Alderton, A., Nelson, R., Harrison, E. M., Sillitoe, J., Bentley, S. D., Barrett, J. C., Estee Torok, M., Goodfellow, I. G., Langford, C., and Kwiatkowski, D. Patterns of within-host genetic diversity in SARS-CoV-2. Aug. 2021.

Turakhia, Y., Thornlow, B., Hinrichs, A. S., De Maio, N., Gozashti, L., Lanfear, R., Haussler, D., and Corbett-Detig, R. Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nat Genet*, 53(6):809–816, Jun 2021.

Viana, R., Moyo, S., Amoako, D. G., Tegally, H., Scheepers, C., Althaus, C. L., Anyaneji, U. J., Bester, P. A., Boni, M. F., Chand, M., Choga, W. T., Colquhoun, R., Davids, M., Deforche, K., Doolabh, D., du Plessis, L., Engelbrecht, S., Everatt, J., Giandhari, J., Giovanetti, M., Hardie, D., Hill, V., Hsiao, N. Y., Iranzadeh, A., Ismail, A., Joseph, C., Joseph, R., Koopile, L., Kosakovsky Pond, S. L., Kraemer, M. U. G., Kuate-Lere, L., Laguda-Akingba, O., Lesetedi-Mafoko, O., Lessells, R. J., Lockman, S., Lucaci, A. G., Maharaj, A., Mahlangu, B., Maponga, T., Mahlakwane, K., Makatini, Z., Marais, G., Maruapula, D., Masupu, K., Matshaba, M., Mayaphi, S., Mbhele, N., Mbulawa, M. B., Mendes, A., Mlisana, K., Mnguni, A., Mohale, T., Moir, M., Moruisi, K., Mosepele, M., Motsatsi, G., Motswaledi, M. S., Mphoyakgosi, T., Msomi, N., Mwangi, P. N., Naidoo, Y., Ntuli, N., Nyaga, M., Olubayo, L., Pillay, S., Radibe, B., Ramphal, Y., Ramphal, U., San, J. E., Scott, L., Shapiro, R., Singh, L., Smith-Lawrence, P., Stevens, W., Strydom, A., Subramoney, K., Tebeila, N., Tshiabuila, D., Tsui, J., van Wyk, S., Weaver, S., Wibmer, C. K., Wilkinson, E., Wolter, N., Zarebski, A. E., Zuze, B., Goedhals, D., Preiser, W., Treurnicht, F., Venter, M., Williamson, C., Pybus, O. G., Bhiman, J., Glass, A., Martin, D. P., Rambaut, A., Gaseitsiwe, S., von Gottberg, A., and de Oliveira, T. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*, 603(7902):679–686, Mar 2022.

Wirth, W. and Duchene, S. GISAIDR, 2022. URL https://doi.org/10.5281/zenodo.6474693.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Supplementary Information

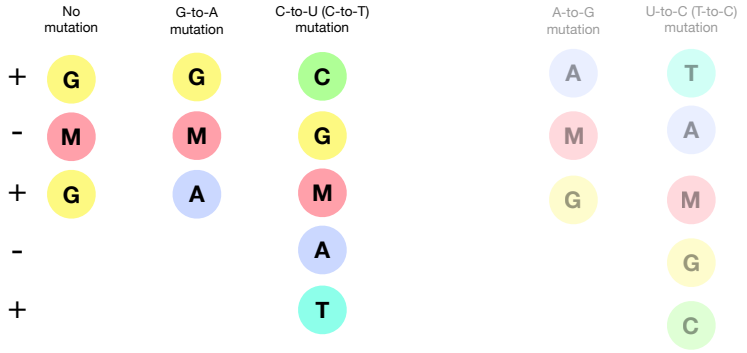

Figure S1. Possible outcomes from MTP incorporation
This figure depicts some of the mutational pathways related to MTP incorporation into MTP. The first column shows what may be a common event, but is not detectable by sequencing. MTP can be incorporated into RNA (pairing with G) and then pair with G again in the next round of synthesis, which will result in **no mutation** in the final sequence. However if the MTP takes on an alternative tautomeric form after incorporation it can bind to A, creating a **G-to-A mutation**. The third column shows that if the positive-sense base is C, then this will bind to a G in the formation of the negative-sense genome. In subsequent replication this negative sense genome can undergo the same G-to-A mutation seen in the second column, which ultimately results in a positive sense **C-to-T** mutation. Although the biases of tautomeric forms for the free and incorporated MTP nucleotides appear to favour these directionalities of mutations, the reverse is also possible, resulting in **A-to-G** and **T-to-C** mutations.

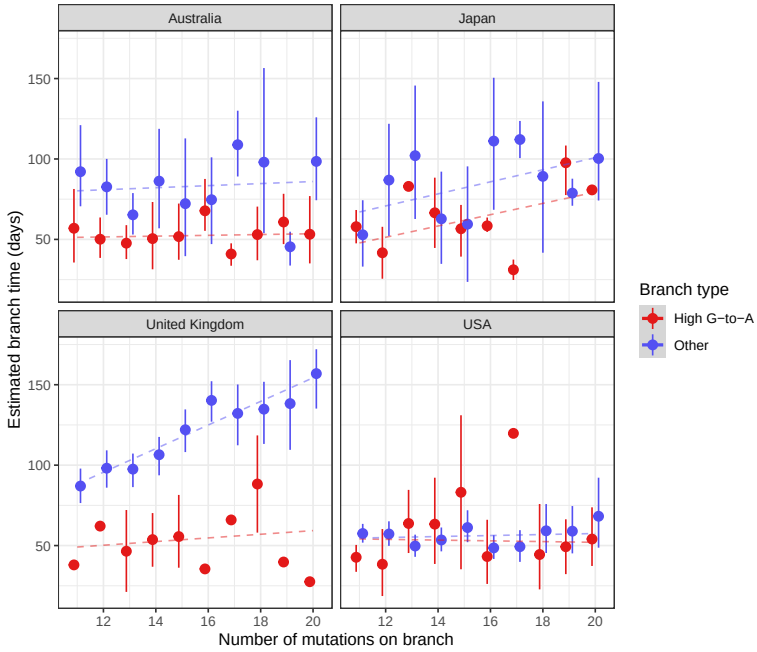

Figure S2. High G-to-A branches involve the same number of mutations occurring in a shorter period of time than other branch types
We used Chronumental to assign a date to all nodes in the global UShER mutation annotated tree, then isolated branch lengths >10 and <=20, and plotted the mean length of time and its 95% confidence interval by number of mutations, as well as a fitted linear model, for Australia, Japan, the UK and the USA. This analysis should be seen only as semi-quantitative due to the fact that the algorithm will attempt to reconcile its chronology with the overall SARS-CoV-2 mutation rate and therefore be conservative, and that sequencing errors and recombination can also drive a large apparent rate of mutations in a short time.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

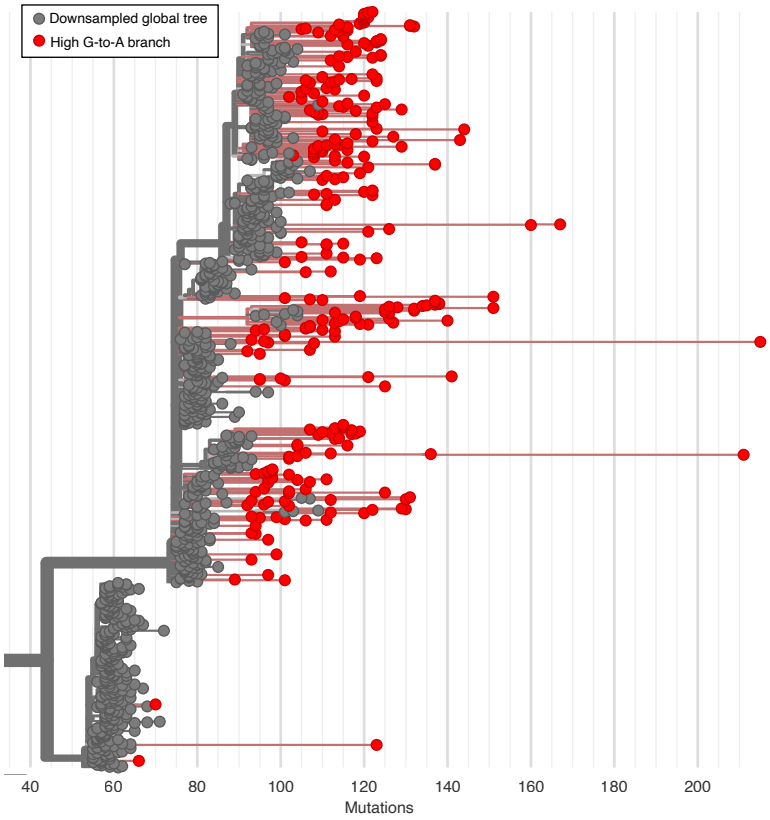

Figure S3. High G-to-A branches from Australia visualised on a downsampled global tree
A complete MAT from Australian sequences was made, and genomes descended from high G-to-A events were identfied. Usher.bio was then used to create a tree with these sequences highlighted on a downsampled global tree, with visualisation performed using Nextstrain.

Figure S4. Distribution of major SARS-CoV-2 variants between placebo and molnupiravir treatments in the AGILE trial dataset.
The proportion of patients infected with each variant is shown. The proportions are similar suggesting that differences between placebo and molnupiravir spectra will not be influenced by previously observed spectrum differences between variants (Ruis et al., Bloom et al.). VOC = variant of concern.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

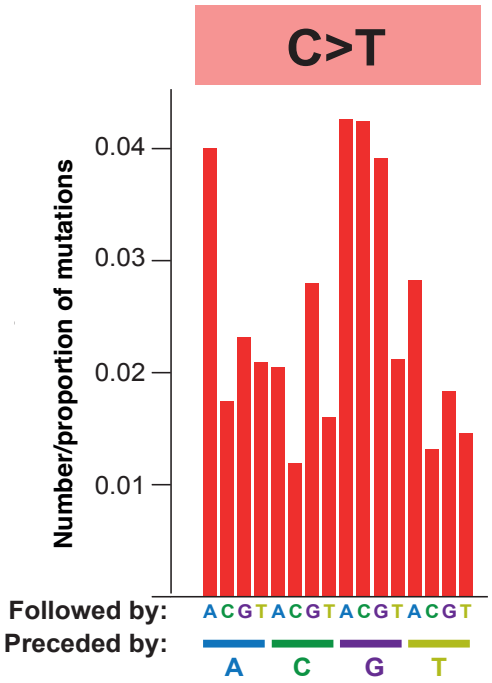

Figure S5. Context locations within the mutational spectrum.
The RNA mutational spectrum contains 12 mutation types, for example C-to-T. The spectrum also captures the nucleotides surrounding each mutation. There are four potential upstream nucleotides and four potential downstream nucleotides. This figure shows the location of each of the 16 contexts within an example mutation type. For example, the leftmost bar represents C-to-T mutations in the ACA context while the second leftmost bar represents C-to-T mutations in the ACC context.

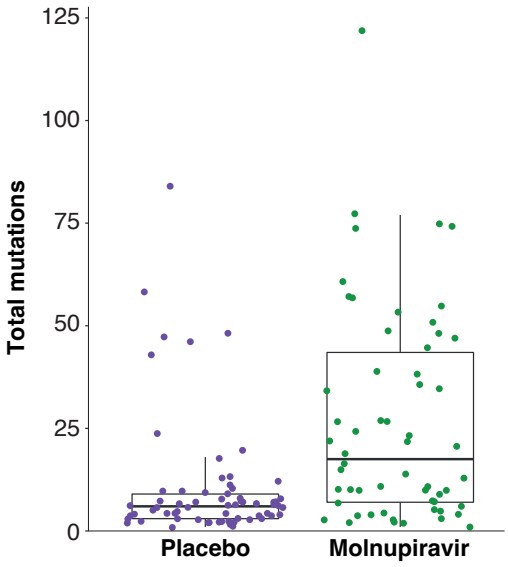

Figure S6. Comparison of the mutational burden in patients treated with placebo and patients treated with molnupiravir.
The total number of mutations across all substitution classes is plotted in the day 5 sample of each patient in the AGILE trial.

medRxiv preprint doi: https://doi.org/10.1101/2023.01.26.23284998; this version posted January 27, 2023. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

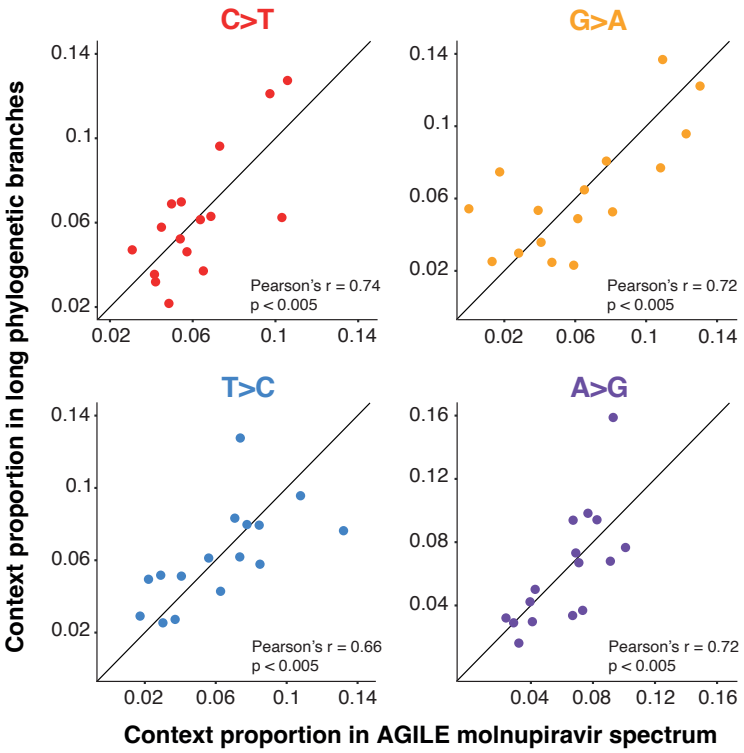

Figure S7. Comparison of contextual patterns within transition mutations between the molnupiravir spectrum calculated from AGILE trial data and the spectrum of long phylogenetic branches.
The proportion of mutations within the respective transition (for example the proportion of C>T mutations that are in the ACA context) is shown. P-values represent the proportion of context randomisations with a Pearson's r correlation at least as great as with the real data (see methods). Proportions are normalised for the number of times the context occurs in the genome.

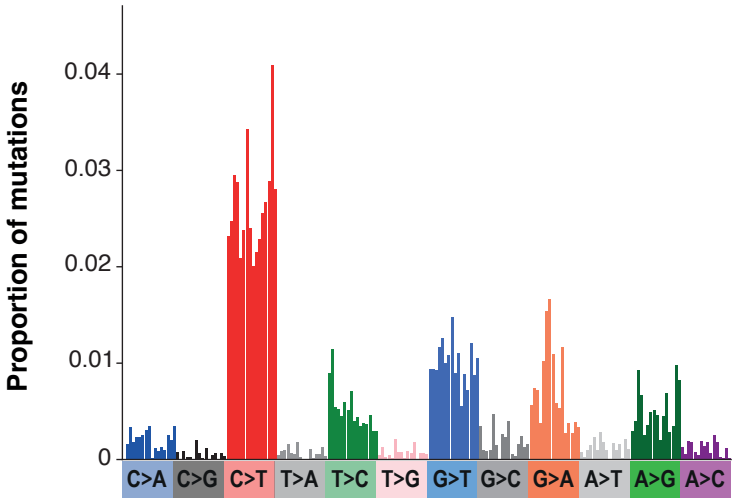

Figure S8. Mutational spectrum of mutations acquired within untreated patients.
Mutations are from Tonkin-Hill et al. (2021).