Applying Site-Saturation Variant Libraries to Stay Ahead of Viral Evolution

ABSTRACT

Viral surveillance is a vital public health activity. Though viral genome sequencing has become more routine, connecting genetic changes to phenotypic outcomes remains challenging. Site-saturation mutagenesis screens can highlight genotype-phenotype associations, but they have historically suffered from practical limitations associated with library generation. This application note explores how Twist's Site-Saturation Variant Libraries (SSVL) can accelerate this critical viral research activity, discussing a specific example where an SSVL enables the investigation of SARS-CoV-2 viral variant evolution.

> "Our ability to see genetic variation outpaces our capacity to evaluate its functional consequences."
>
> Tyler Starr, Ph.D., Assistant Professor of Biochemistry, U. of Utah.

THE HASTENING PACE OF VIRAL SURVEILLANCE

Viral surveillance is the systematic collection and analysis of data related to the occurrence and spread of viral pathogens to dull the impact of emerging outbreaks and pandemics. By tracking viral evolution and novel variants, viral surveillance programs collectively serve as an early warning system that provides governments, public health organizations, diagnostic developers, and pharmaceutical companies with valuable lead time, such that they can formulate public health and therapeutic intervention plans.

Genomic pathogen surveillance depends on global collaboration between a network of collection sites and sequencing nodes that take patient samples and generate genomic data.[1] Researchers can interrogate that data for genetic differences that may explain or predict pathogenic outbreaks. Genetic variants that become fixed in pathogenic populations and correlate with clear phenotypic changes, like elevated transmissibility, lethality, and unique symptoms are of particular interest.

As a relevant example, the COVID-19 pandemic created a mass impetus to improve the speed, volume, and quality of viral surveillance efforts.[2] This groundswell led to the build-up of global surveillance capacity, enabling public sharing of viral sequences at unprecedented levels and near-real-time monitoring of SARS-CoV-2 variant emergence.[3]

However, the COVID-19 pandemic also exposed weaknesses in existing surveillance programs. Chiefly, "our ability to see genetic variation outpaces our capacity to evaluate its functional consequences," says Tyler Starr, Ph.D., Assistant Professor of Biochemistry at the University of Utah, whose lab studies molecular evolution in viruses and immunity. In a bid to speed up studies on the functional consequences of viral evolution, researchers like Starr have turned to site-saturation mutagenesis.

SITE-SATURATION MUTAGENESIS

Rather than wait for a mutation to occur, viral researchers can perform site-saturation mutagenesis studies to anticipate possible mutants and study their functional impact ahead of their detection in the wild.

Site-saturation mutagenesis systematically alters each codon position of a gene to create libraries wherein every natural amino acid is individually replaced by all possible amino acid alternatives (**Figure 1**). With these viral gene libraries, researchers can create a vast array of protein variants that each contain a single mutation within the protein sequence. By cloning and expressing individual mutant viral proteins within a multiplexed functional assay, researchers can safely study viral evolution and investigate its impact on transmission, pathogenesis, and immune escape without generating live viruses.

THE CHALLENGE OF TRADITIONAL LIBRARY GENERATION STRATEGIES

Historically, researchers generated site-saturation mutagenesis libraries using traditional PCR-based methods, like site-specific mutation using degenerate primers or error-prone PCR (epPCR).[4]

However, since these approaches rely on biological interactions and enzymatic activities, it can be difficult to achieve complete coverage and high uniformity of each possible codon variant.[5] PCR approaches can also introduce additional mutations due to polymerase errors, especially when using non-proofreading polymerases in epPCR. Because of these limitations, PCR-based methods often force researchers to oversample mutations, wasting considerable in the process. In response, they must compromise on the scale and breadth of their study, narrowing the number of mutagenesis sites tested to reduce overall costs. Such a compromise may come at the expense of public health as important variants are overshadowed or left unexamined.

TWIST'S SITE-SATURATION VARIANT LIBRARIES (SSVL)

To carry out truly comprehensive mutagenesis studies, researchers can forgo traditional in-house, PCR-based methods, opting instead to use Twist's Site-Saturation Variant Libraries (SSVLs). Instead of relying on error-prone polymerases to generate variant libraries, SSVLs leverage phosphoramidite chemistry and Twist's proprietary silicon support system to directly synthesize variant DNA sequences with high precision at an unprecedented scale.[6] With highly uniform synthesis, Twist's synthetic libraries give researchers control over the precise composition of the mutagenesis library, reducing the need for oversampling and empowering more comprehensive viral variant screening. In this way, Twist's SSVLs allow researchers to keep pace with modern viral surveilance efforts, annotating the functional consequences of viral variants at scale.

Critically, Twist's platform is built for both speed and accuracy. Twist is able to deliver high-quality SSVLs with multiple genetic backgrounds in just 3-4 weeks in pooled or arrayed formats to enable versatile downstream applications. In turn, researchers save time on library generation and can instead focus on building predictive phenotypic assays. Given the importance of speed in viral surveillance, the SSVL's rapid synthetic assembly provides a major advantage for quickly getting critical data to researchers and public health decision-makers.

This application note explores how a team led by Tyler Starr, Ph.D., applied SSVL to study the impact of genetic epistasis in SARS-CoV-2 variant evolution.[7] More specifically, the study sought to understand how accumulated mutations in the Spike protein receptor-binding domain (RBD) interact with other mutations.

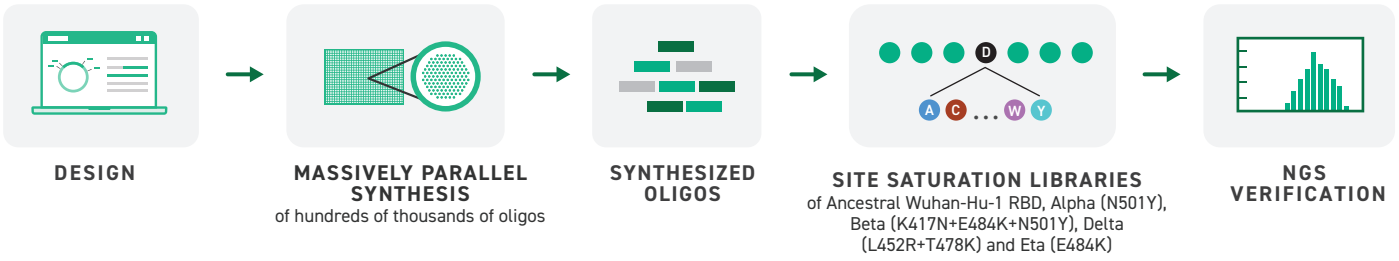

Figure 1: Generation of Site-Saturation Variant Libraries in Multiple Variant Genetic Backgrounds with Twist.

RESULTS AND WORKFLOW

SARS-CoV-2 viral surveillance has paid particular attention to tracking changes in the Spike protein's receptor binding domain (RBD), given its importance in ACE2 receptor binding and vaccine and monoclonal antibody therapeutic development. In a previous study,[8] Starr and collaborators generated an in-house mutagenesis library of the original Wuhan strain RBD, ultimately identifying the N501Y mutation as a boon to ACE2 binding. In time, this mutation was enriched in the Alpha, Beta, and Gamma variants. This means that the mutation was detected before it evolved in nature, indicating that mutagenesis studies could help predict evolutionary steps.

Starr and his collaborators sought to continue this work by studying how specific SARS-CoV-2 mutations influence others and guide viral evolution. In other words, the team wanted to explore SARS-CoV-2 RBD epistasis and how individual mutational effects vary across different viral variants, including Wuhan-Hu-1, Alpha, Beta, Delta, and Eta.[7] In doing so, researchers may gain a deeper understanding of SARS-CoV-2 evolution and the threat that different variant combinations can pose to public health.

Library Variant Coverage and Quality

SSVLs were ordered to cover each of the five viral variant backgrounds (19 amino acid substitutions at 201 positions for each) to study how mutational impact differs across RBD variants. After synthesis, the SSVL deep sequencing QC reports generated by Twist showed good uniformity and high site representation across all genetic backgrounds. SSVL provided complete or near-complete coverage and even mutational uniformity in all cases.

After covering the few remaining positions with in-house mutagenesis, researchers then barcoded the mutant libraries and cloned them into a vector backbone for a phenotypic yeast display assay. Long-read sequencing was used to independently evaluate the library quality and establish annotated barcodes for specific mutations.

In another SARS-CoV-2 study, a team led by Starr[9] similarly applied Twist SSVL libraries and site saturation mutagenesis to investigate Omicron variant evolution. The authors developed and applied an improved method for appending molecular barcodes to the SSVL libraries using direct three-fragment Gibson assembly to minimize the mutational spreading that can occur during barcoding with low-cycle PCR. In this report, authors sequenced over 200,000 variants for both Omicron BA.1 and BA.2 libraries to evaluate quality (**Figure 2**).

Virtually all variants were found to have a single mutation (**Figure 2A**), indicating high SSVL accuracy (**Figure 2B**). This is critical for ensuring assay results can be attributed to a single genetic change. Plotting codon mutation frequency (**Figure 2C**), it was also apparent that SSVL provided steady uniformity and complete coverage, as evidenced by low variation across all codon positions. In addition, virtually no stop or synonymous mutations were present, with the exception of a small sampling used as controls.

Combined with Twist's own sequencing reports, this data demonstrates that SSVL provides precise single mutant variants and balanced mutation rates across all positions. The QC data shows good coverage at each position without unnecessary stop codons or synonymous mutation, indicating high library efficiency. Such high-quality SSVL synthesis gives researchers the opportunity for more efficient, large-scale variant screening. Rather than compromise on the scale of variant screening and potentially overlook important variants, SSVLs allow researchers to pursue more comprehensive screening without sacrificing efficiency or accuracy.

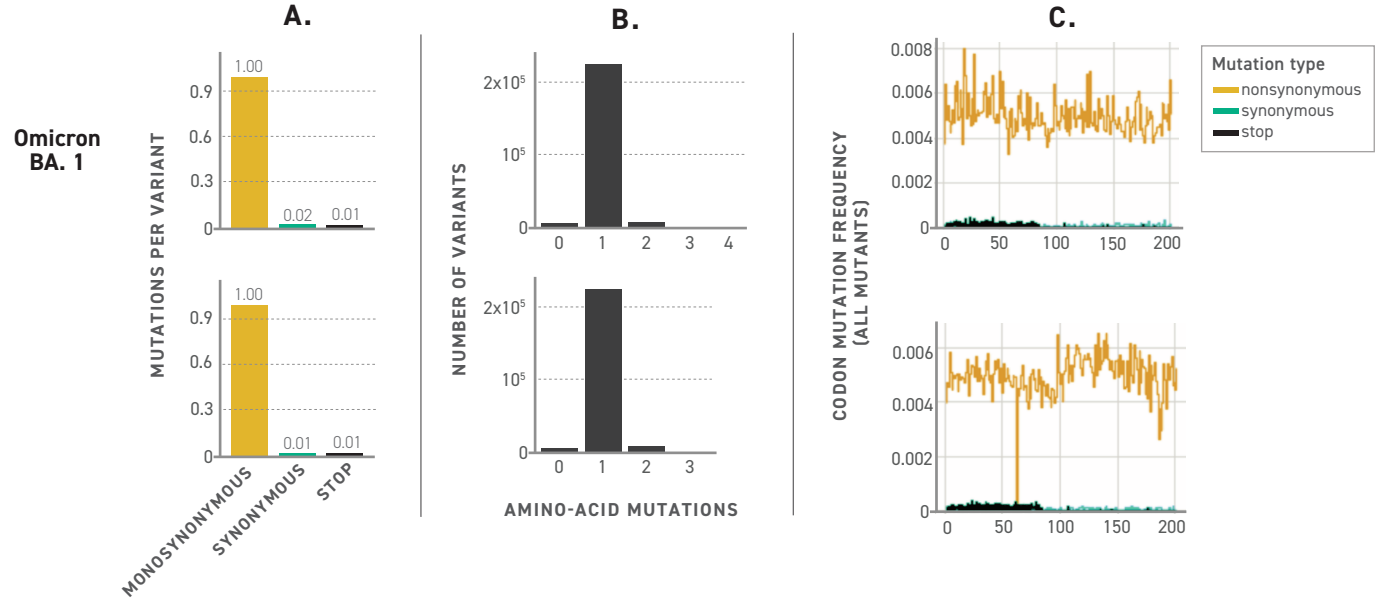

Figure 2: Long-Read Sequencing Confirms SSVL Quality by Amino Acid Mutation Count and Codon Mutation Frequency.

Phenotypic Assays Exploring Viral Variant Library

With libraries in hand, Starr and his group set to the task of evaluating how each mutational combination affected the Spike protein's ability to bind to its target (ACE2). Using a yeast display phenotypic assay system, researchers collected thousands of independent binding affinities covering each RBD variant library in a single experiment.

Using the phenotypic assay and site saturation mutagenesis libraries, epistatic relationships in viral variant mutations were explored. Epistatic shifts in mutational effects on ACE2 binding were plotted across each RBD variant and benchmarked to the original Wuhan strain (**Figure 3**). Though a number of shifts were identified, most were attributable to the N501Y mutation, given that the strongest shifts occur in variant backgrounds where N501Y was present (**Alpha and Beta, Figure 3**).

The largest individual shift in mutational effect was observed at the 498 RBD position, where the native amino acid comes into direct contact with the amino acid located at N501. Double mutants containing Q498R and N501Y demonstrated a 25X and 387X affinity improvement over the Beta (N501Y) and Wuhan strains, respectively (**Figure 4A**). Interestingly, when on its own, Q498R actually reduced ACE2 binding, which further demonstrates the importance of evaluating combinatorial mutations and epistatic relationships when tracking viral evolution.

The Q498R mutation was not detected disproportionally in naturally occurring viral variants containing Y501 until its emergence and fixation in Omicron lineages. The authors hypothesized that the substantial affinity gain of the double mutant alone was not advantageous to viral evolution until paired with additional mutations that supply other fitness benefits to Omicron. To explore this, single mutation effects of Omicron RBD mutations were plotted by their individual effect on both the Wuhan and Beta backgrounds (**Figure 4B**). Cumulative single-mutant effects of Omicron RBD mutations in the Wuhan background were predicted to significantly reduce ACE2 affinity (**Figure 4B, left**). In contrast, nearly no affinity change was observed in the Beta background, which includes N501Y (**Figure 4B, right**).

The authors reasoned that N501Y/Q498R may provide an affinity buffer that allows the Omicron spike to tolerate the accumulation of other mutations that decrease ACE2 binding but aid antibody escape. For example, E484 and K417 have known effects on antigenic selection and escape from neutralizing antibodies.[10]

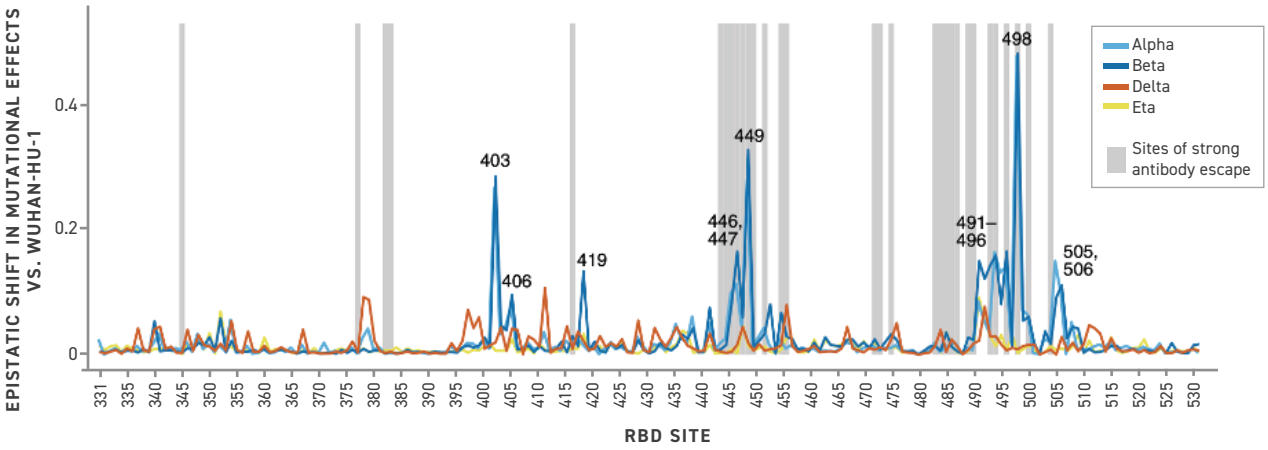

Figure 3: Epistatic Shift in Mutational Effects Compared to the Wuhan Strain.

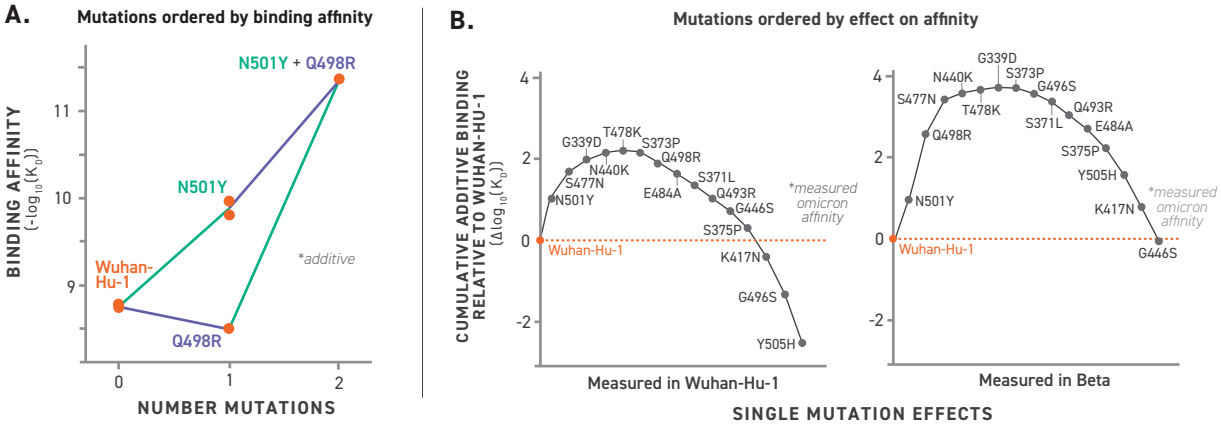

Figure 4: Functional Impact of Variant Mutations and Epistatic Mutations.

CONCLUSION

The results presented in this application note demonstrate how site-saturation mutagenesis can support viral surveillance. Using SSVLs, these studies can shed light on functional changes in viral variants and potential evolutionary paths, even before they occur in nature.

In particular, this study helps explain how the Omicron variant could sustain more antigenic drift to avoid existing and vaccination-associated immunity while maintaining high human transmission. The Starr lab has continued using Twist's SSVL to explore viral evolution, including epistatic drift[11] and antibody recognition[12] in additional SARS-CoV-2 variants.

Since viruses like SARS-CoV-2 evolve quickly and can sample massive evolutionary space, it is important that site-saturation mutagenesis studies can be performed rapidly across multiple circulating variants simultaneously. For this reason, Twist SSVL helps to remove barriers for large-scale studies of pathogen evolution by allowing researchers to get functional results faster, enabling greater experimental speed and consistency.

Rather than spend time generating libraries, researchers can work in parallel with Twist and devote their attention to developing phenotypic assays or other experimental steps. Using SSVL eliminates the responsibility of creating libraries and optimizing associated methods, which, according to Starr, can save "months" of researcher time, especially when a researcher lacks library generation experience.

The reliability, uniformity, and scalability of Twist's SSVL provide viral surveillance researchers with a means to stay ahead of pathogen evolution, closing the disconnect between genomic detection and functional assessment.

At Twist Bioscience, we work in the service of customers who are changing the world for the better. In fields such as medicine, agriculture, industrial chemicals, and data storage, by using our synthetic DNA tools, our customers are developing ways to better lives and improve the sustainability of the planet.

To learn more about the Starr lab's phenotypic assays and how they apply them to study viral evolution, watch this webinar on Unveiling the Dynamics of SARS-CoV-2 Variant Evolution.

REFERENCES

1. WHO Epidemic and Pandemic Preparedness and Prevention, Global Genomic Surveillance, & Research for Health Teams (2022) Global genomic surveillance strategy for pathogens with pandemic and epidemic potential, 2022–2032. WHO. 1-32.

2. Carroll D, Morzaria S, Briand S, Johnson C K, Morens D, Sumption K, Tomori O, & Wacharphaueasadee S (2021) Preventing the next pandemic: the power of a global viral surveillance network. BMJ, 372, n485.

3. Attar Cohen H, Mesfin S, Ikejezie J, Kassamali Z, Campbell F, Adele S, Guinko N, Idoko F, Mirembe BB, Mitri ME, Nezu I, Shimizu K, Ngongheh AB, Sklenovska N, Gumede N, Mosha FS, Mohamed B, Corpuz A, Pebody R, Marklewitz M, Gresh L, Mendez Rico JA, Hundal K, Kato M, Babu A, Archer BN, le Polain de Waroux O, Van Kerkhove MD, Mahamud A, Subissi L, Pavlin BI (2023) Surveillance for variants of SARS-CoV-2 to inform risk assessments. Bull World Health Organ. 101, 707-716.

4. Alejaldre L, Pelletier JN, & Quaglia D (2021) Methods for enzyme library creation: Which one will you choose? BioEssays, 43, 2100052.

5. Acevedo-Rocha CG, Reetz MT, & Nov Y (2015) Economical analysis of saturation mutagenesis experiments. Scientific Reports, 5, 10654.

6. Twist Bioscience (2018). A Simple Guide to Phosphoramidite Chemistry and How it Fits in Twist Bioscience's Commercial Engine. https://www.twistbioscience.com/blog/science/simple-guide-phosphoramidite-chemistry-and-how-it-fits-twist-biosciences-commercial

7. Starr TN, Greaney AJ, Hannon WW, Loes AN, Hauser K, Dillen JR, Ferri E, Farrell AG, Dadonaite B, McCallum M, Matreyek KA, Corti D, Snell G, Bloom J (2022) Shifting mutational constraints in the SARS-CoV-2 receptor-binding domain during viral evolution. Science, 377, 420-424.

8. Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesler D, & Bloom JD (2020) Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell, 182 (5), 1295-1310.

9. Starr TN, Greaney AJ, Stewart C M, Walls AC, Hannon WW, Veesler D, Bloom JD (2022) Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. PLoS Pathogens, 18 (11), e1010951.

10. Greaney AJ, Starr TN, Barnes CO, Weisblum Y, Schmidt F, Caskey M, Gaebler C, Cho A, Agudelo M, Finkin S, Wang Z, Poston D, Muecksch F, Hatziioannou T, Bieniasz PD, Robbiani DF, Nussenzweig MC, Bjorkman PJ, & Bloom JD (2021) Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nature Communications, 12, 4196.

11. Taylor AL & Starr TN (2023) Deep mutational scans of XBB.1.5 and BQ.1.1 reveal ongoing epistatic drift during SARS-CoV-2 evolution. PLoS Pathogens, 19 (12), e1011901.

12. Ruiz F, Foreman W, Lilly M, Baharani VA, Depierreux DM, Chohan V, Taylor AL, Guenthoer Jamie, Ralph D, Matsen IV FA, Chu HY, Bieniasz PD, Côté M, Starr TN, & Overbaugh J. (2024) Delineating the functional activity of antibodies with cross-reactivity to SARS-CoV-2, SARS-CoV-1 and related sarbecoviruses bioRxiv 2024.04.24.590836

DOC-001507 REV 1.0