PROCEEDINGS OF THE ROYAL SOCIETY B

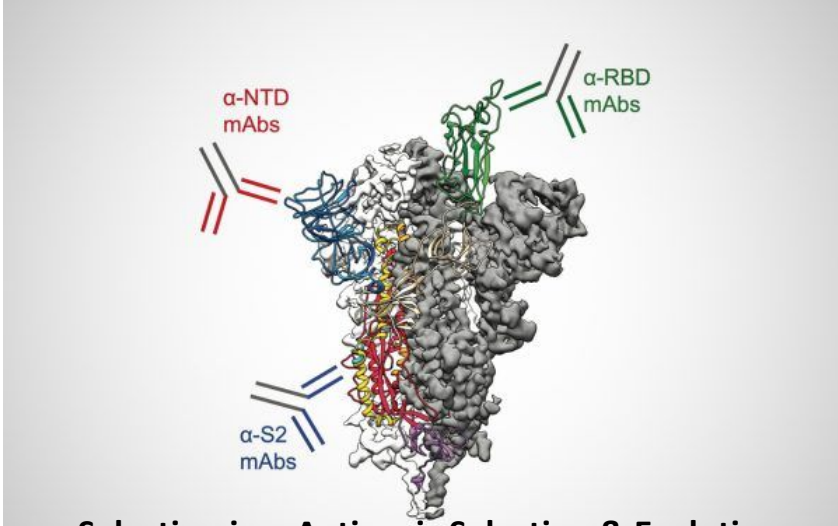
*Proc. R. Soc. B* (2012) **279**, 1249–1258
doi:10.1098/rspb.2011.2293
*Published online* 4 January 2012

*Review*

# The role of robustness in phenotypic adaptation and innovation

**Andreas Wagner***

**Heritable Variation**




**Selection,i.e., Antigenic Selection & Evolution**

- Central question
  - Why doesn't *mutational robustness* hinder adaptation?
    - Mutational robustness is resistance to mutational perturbations
    - Here he means single point mutations in the genotype that don't change the phenotype
    - The direct effect is to hinder adaptation because neutral mutations make it hard to change the phenotype
- The 2 part answer
  1. Robustness causes genotype networks to emerge
     - genotype networks are large connected sets of genotypes with the same phenotype (and therefore the same fitness)
  2. The evolutionary dynamics of evolving populations on genotype networks increase the ability *of the population* to adapt

PROCEEDINGS OF THE ROYAL SOCIETY B

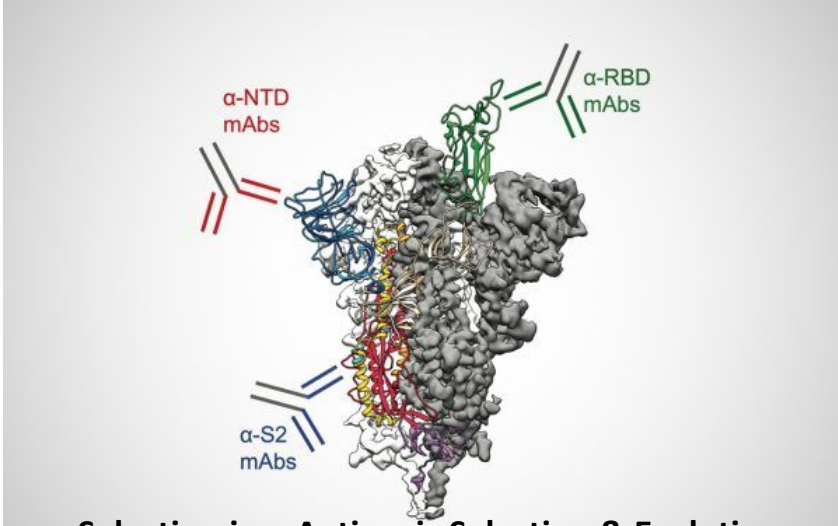
*Proc. R. Soc. B* (2012) **279**, 1249–1258
doi:10.1098/rspb.2011.2293
*Published online* 4 January 2012

*Review*

# The role of robustness in phenotypic adaptation and innovation

**Andreas Wagner***

**Heritable Variation**




**Selection,i.e., Antigenic Selection & Evolution**

- Central question
  - Why doesn't *mutational robustness* hinder adaptation?
    - Mutational robustness is resistance to mutational perturbations
    - Here he means single point mutations in the genotype that don't change the phenotype
    - The direct effect is to hinder adaptation because neutral mutations make it hard to change the phenotype
- The 2 part answer
  1. Robustness causes genotype networks to emerge
     - genotype networks are large connected sets of genotypes with the same phenotype (and therefore the same fitness)
  2. The evolutionary dynamics of evolving populations on genotype networks increase the ability *of the population* to adapt
- The phenotype of an organism is robust if it persists when perturbed
  - Think of phenotype as the shape of a protein which causes, for example, the color of a flower or the binding of a protein segment (antigen) to an antibody.
  - Perturbations include
    - Environmental, like temperature change or a new antibody has evolved
    - Mutation, i.e. a base pair change that makes the leaves wilt or the color change or evades the antibody or changes the ability to bind to a cell
- Most mutations are detrimental
  - Organisms have evolved to be robust to mutations: the phenotype usually doesn't change with one mutation in the genotype (bitflips have no effect, errors are corrected)
  - But phenotypic variation is essential to evolution (what are the key 3 ingredients for evolution)?

# Evolutionary Conflict?

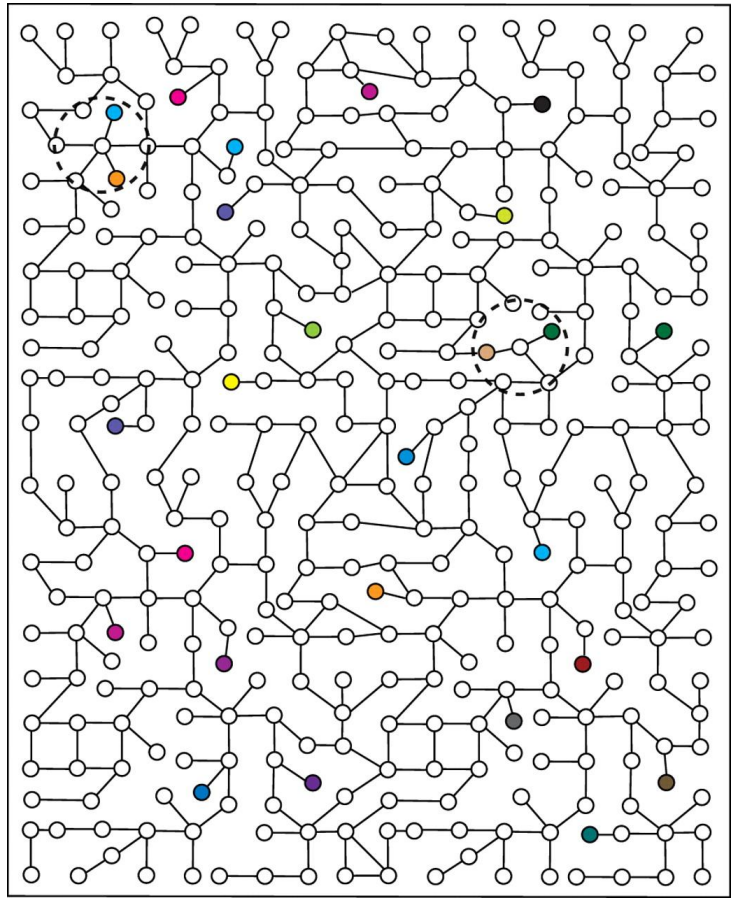
- Conflict?:
  - Phenotypic variability and robustness may seem to oppose each other
  - Because in a robust system, mutations do not easily change a phenotype

- But it's so much more complex than that!

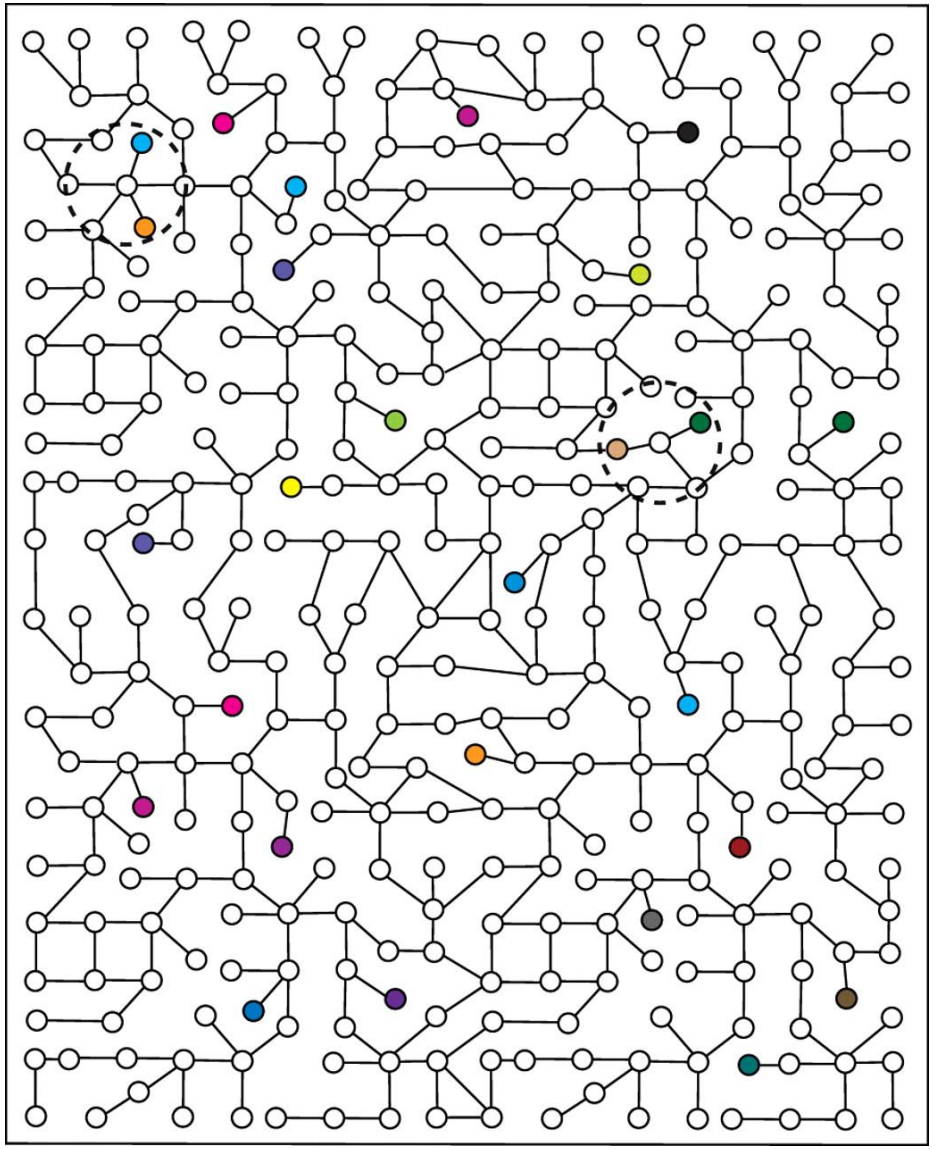
- Let's investigate the relationship between genotypic change and phenotypic change
  - The conflict is resolved by evolutionary dynamics (at a higher level of organization)
  - Evolutionary conflict the cost/benefit of phenotypic variation between an individual and population
    - Which level benefits from variation (individual or population)
    - (Is use of a neutral network a form of group selection ?)

- Evolution explores the new (progressive) while saving the old (conservative)

- Genotype space:
  - 20 (adenine, tryptophan…) amino acids so there are $20^{L}$ protein genotypes where L is the length of the amino acid string
  - 4 base pairs (A,C,G, T) (or substitute U) that make a DNA (or RNA sequence) so $4^{3L}$
    - Note: there are 3 times more base pairs for a given amino acid string because 3 base pairs form 1 codon encoding 1 amino acid
- Here assume the phenotype is the shape of a protein (which determines binding affinity of a protein to an antibody) encoded by a string of base pairs that encode a string of amino acids
- Think of a population as a collection of genotypes. **Evolution acts on this population and the "fittest" individuals survive and reproduce**
- Nodes in the graphs are individuals with a particular genotype. Edges connect individuals that vary by one point mutation (in a base pair A,C,G or T).
  - In the graph different colors are different phenotypes



- How to find a better genotype?
  - For this particular environment, when most mutations are detrimental or lethal?
- Many genotypes produce the same phenotype
  - Example: Hemoglobin (oxygen carrying red blood cell)
    - 95% of genes can be changed & same function is retained!!!
  - Compensation
    - slightly bad mutations can be compensated for so phenotype *appears* unaffected against some stated benchmark or metric
  - Can traverse the network with no detrimental impact on phenotype (the network is NEUTRAL if all individuals share the “same” phenotype which have equal fitness in the given environment)



Wagner A Proc. R. Soc. B 2012;279:1249-1258

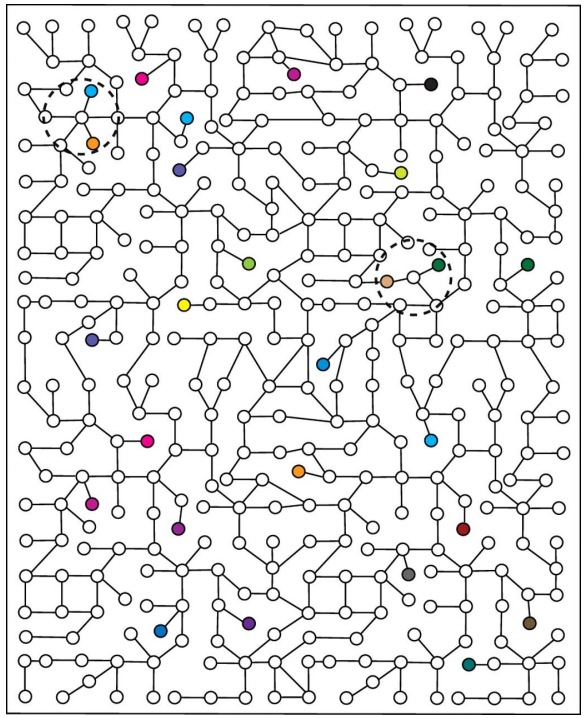
- The neighborhood:
  - Phenotypes in the neighborhood can be reached by mutation
  - The size of the neighborhood measures how phenotypically variable a particular genotype is
    - i.e. if you have many kids, each with a random mutation, how phenotypically different will they be?
  - The neighborhoods of different genotypes contain different phenotypes

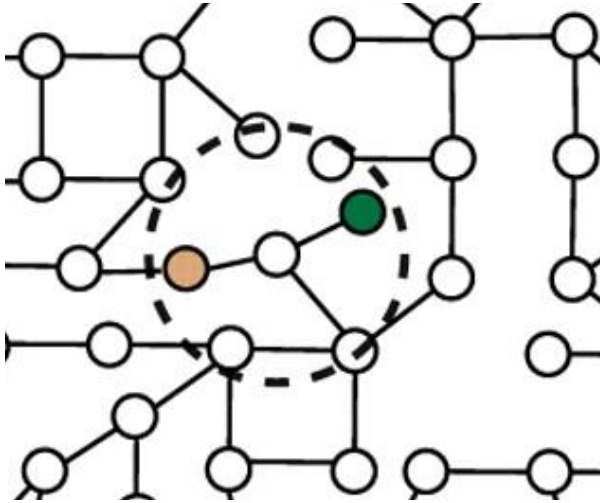
- 1) genotype networks, allow individuals in a population to preserve their phenotype while changing their genotype in many small steps that, cumulatively, can add up to substantial divergence.
- 2) the different genotypes on a genotype network can explore different phenotypes, because their neighborhoods contain different novel phenotypes.

- (Implicitly assumed the simplification that one genotype has one phenotype.  IS this true?)

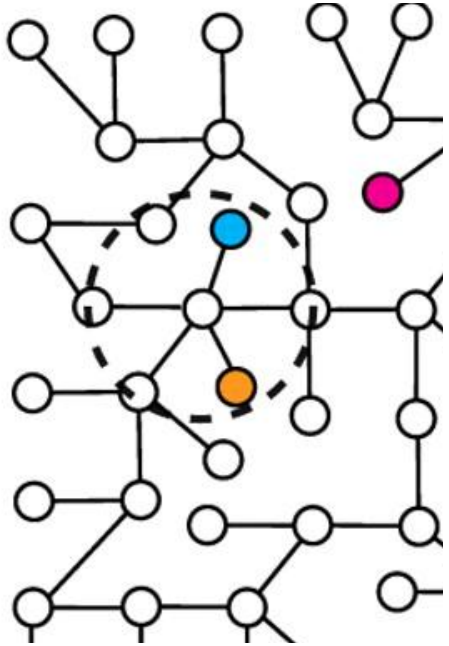
# Conflict resolved: Robustness reduces phenotypic variability at the local level, but not the population level

- A genotype is more robust to mutation the more neighbors it has with the same phenotype (P) as itself





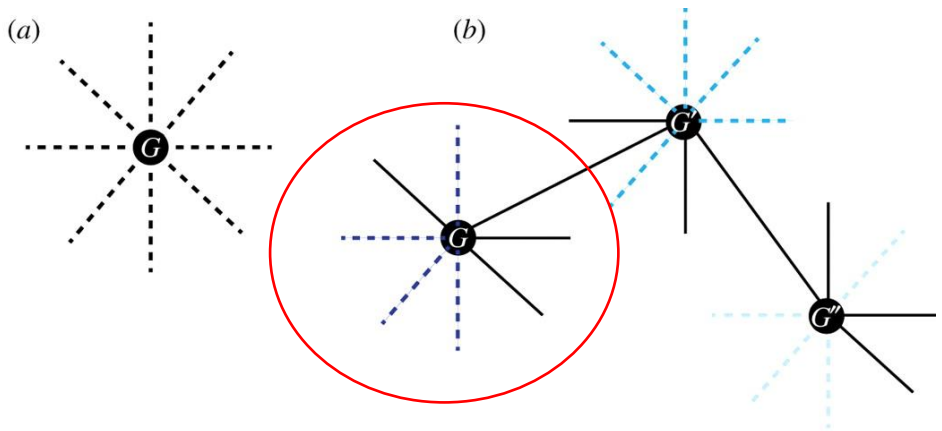
Less robust: 66% of mutations
Lead to new phenotypes



More robust: 50% of mutations lead to the same open circle phenotype

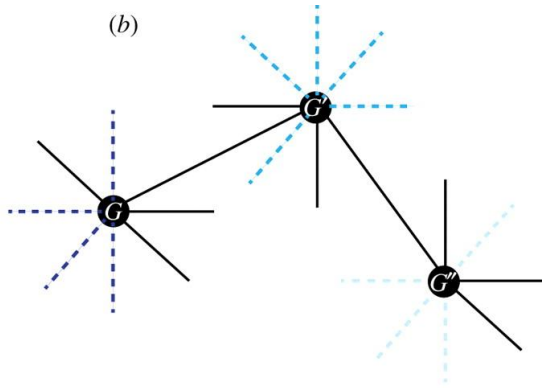
# Conflict resolved: Robustness reduces phenotypic variability at the local level, but not the population level

- A genotype is more robust to mutation the more neighbors it has with the same phenotype (P) as itself
- Compare G1 with no neighbors with the same phenotype to G2 with ½ its neighbors having the same phenotype

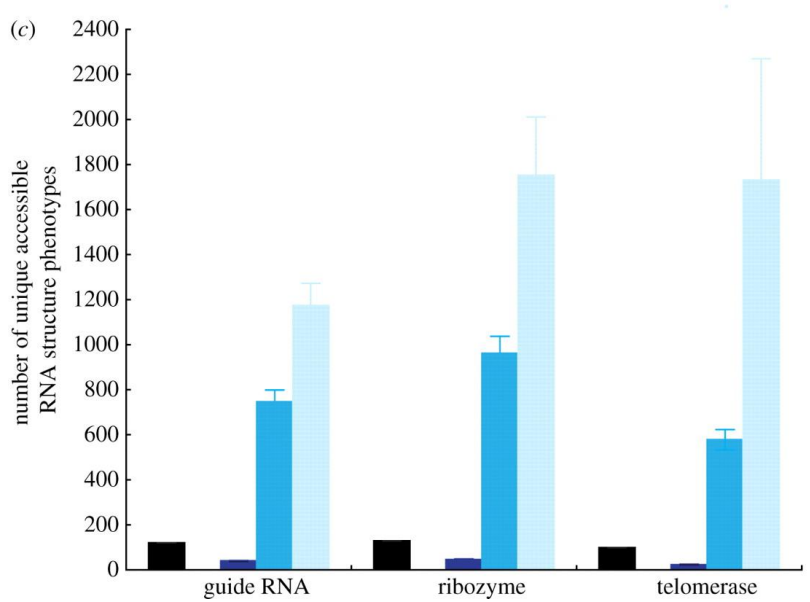



(a): not robust, phenotypically more variable
(b): G is 50% robust & phenotypically less variable
(dashed lines connect to a different phenotype)

# But it’s more complex than that




- G has neighbor G’ who is the same phenotype
- G’ has neighbor G’’ who is also the same phenotype
  - G, G’ and G’’ can mutate back and forth with no change in phenotype
  - When the environment changes, they have mutational access to all neighbors of all 3 genotypes
  - This is backed up by experiments



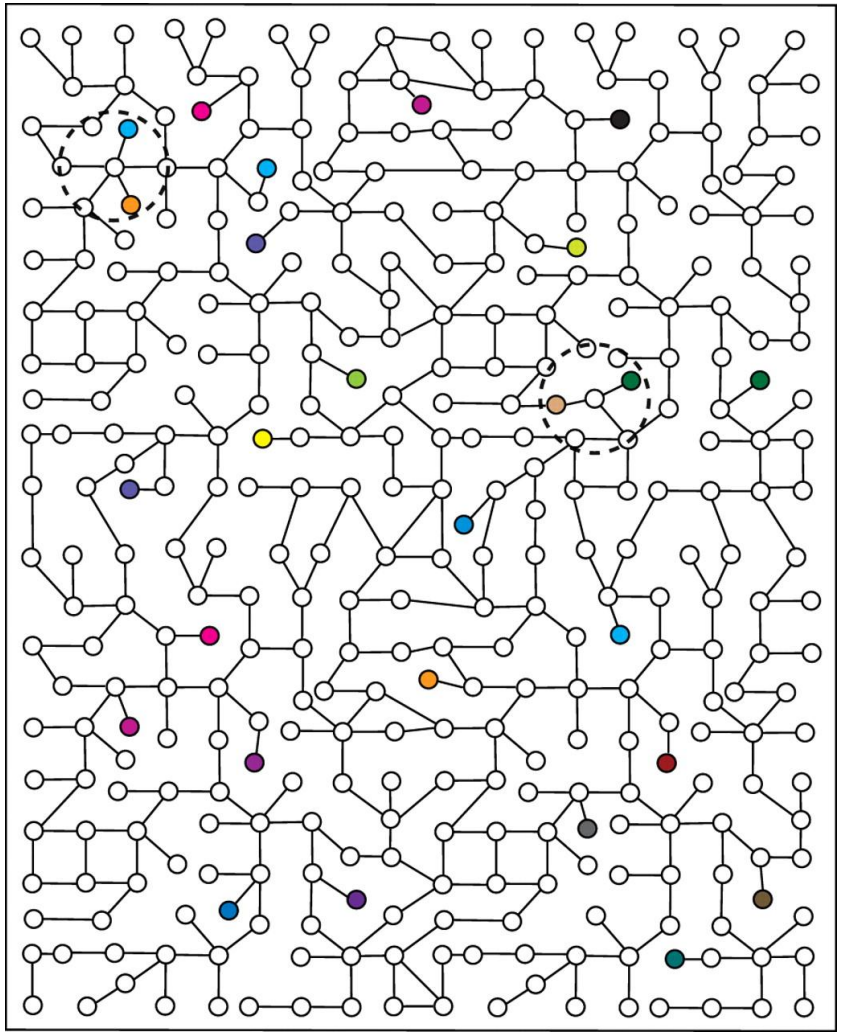
Calculations: there are only 7.2 x10^3 genotypes that differ from the original RNA with L = 40

ACUGGUACAAAACCCCUUUUGGGGAAAACCCCUUUUGGGG

120 ways to have 1 mutation (40 letters, each can change to 3 others)
Each of those has 120 ways to mutate (14,400 total mutations, ½ are repeats)

There are only a small number of mutations from any individual, but MANY mutations in a neighborhood with large radius

## Connected genotype networks facilitate accessibility of diverse phenotypes.

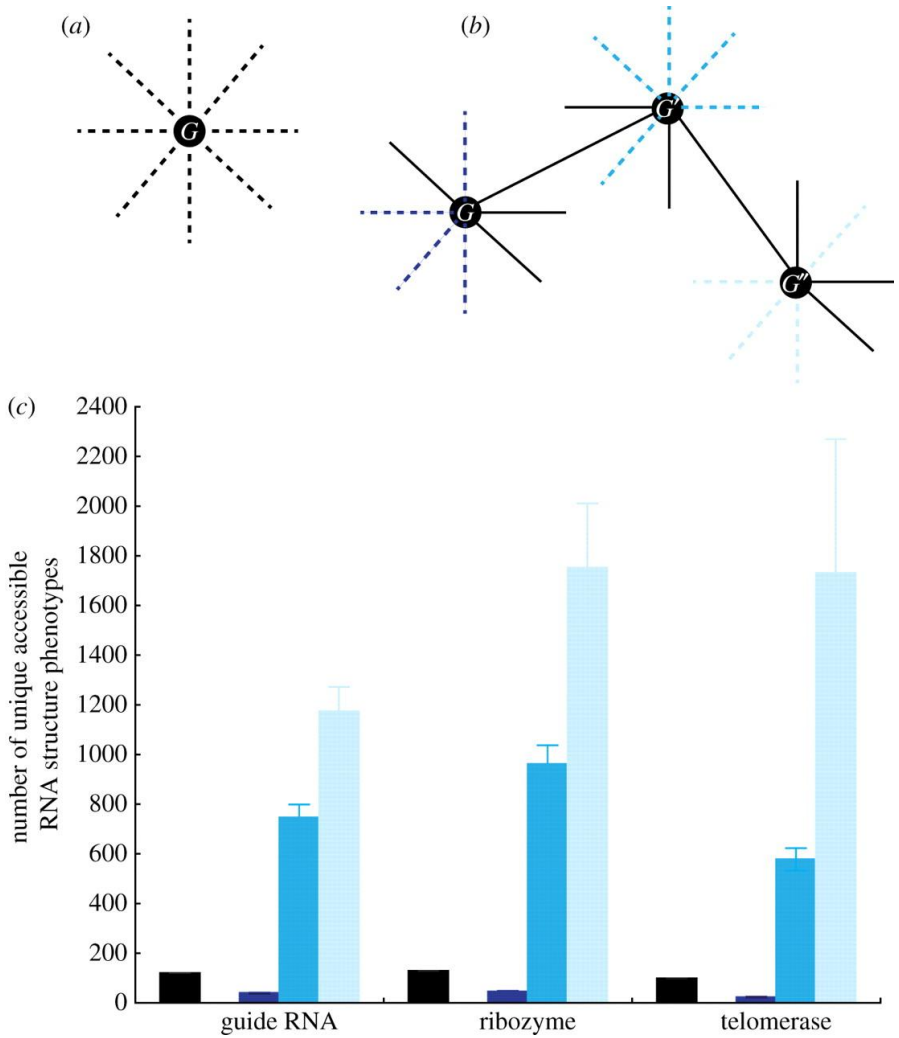


Wagner A Proc. R. Soc. B 2012;279:1249-1258

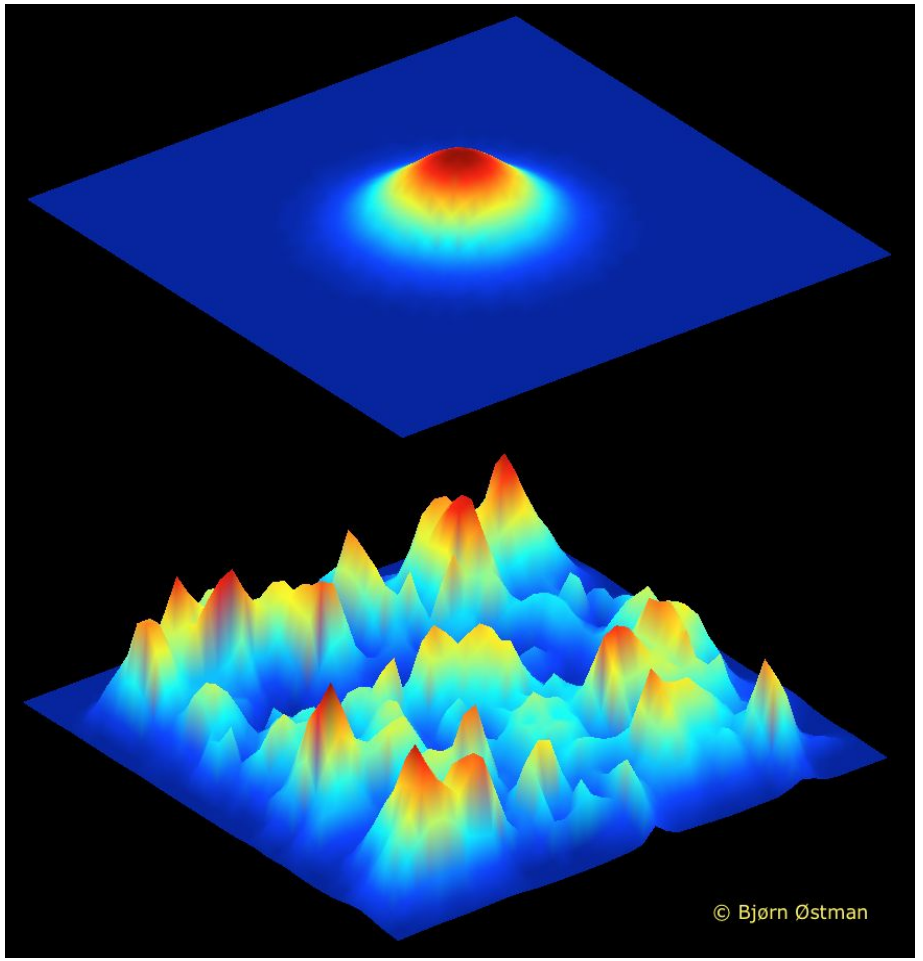


PROCEEDINGS OF THE ROYAL SOCIETY B

# Robustness makes many phenotypic variants accessible to mutations.

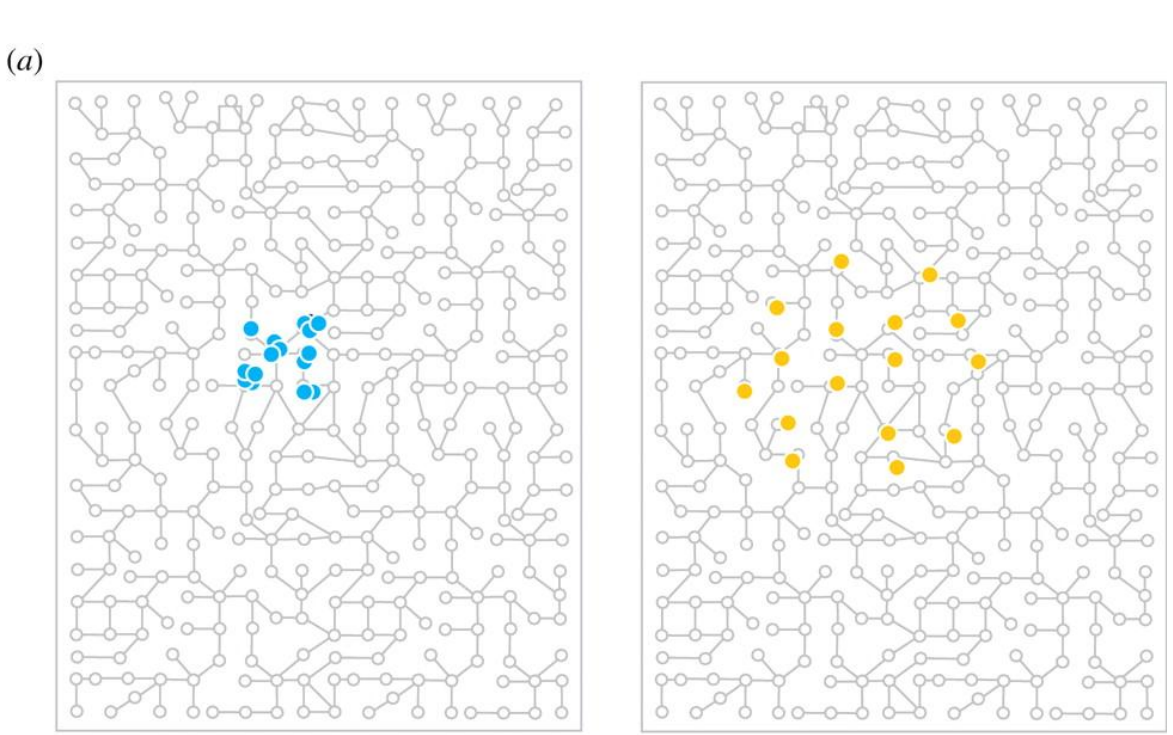


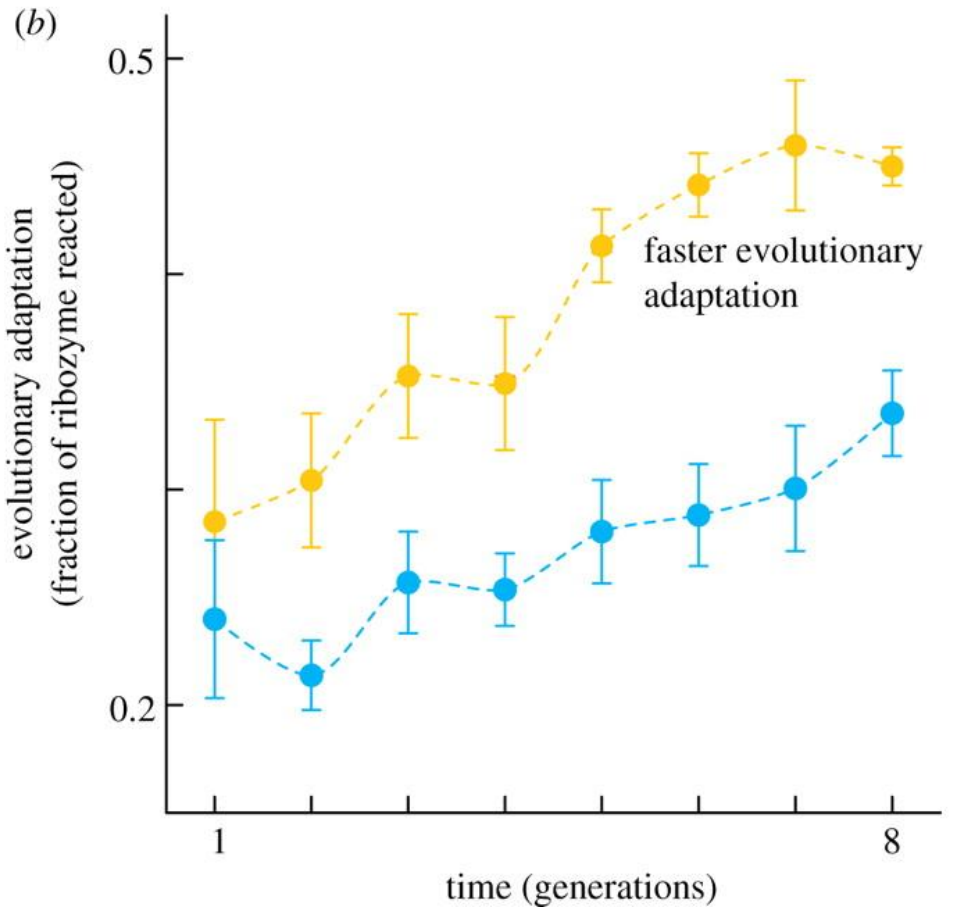

Wagner A Proc. R. Soc. B 2012;279:1249-1258




http://io9.com/these-sweet-3d-fitness-landscapes-show-evolution-at-wor-1563440239

**Cryptic variation can facilitate evolutionary adaptation.**




Wagner A Proc. R. Soc. B 2012;279:1249-1258

Wagner's Neutral Network paper proposed a question: Given an initial string 40 nucleotides long, how many neighbors does the string have in the mutational network?

Consider this string of length 40:
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
Call this initial string G0.

How many genomes are 1 mutation away (we'll call this G1)?
40 x 3 = 120 (each nucleotide can be changed to 3 others.) Here are 6 of the strings in G1.

CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

ACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
ATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
…

How many strings are 2 mutations away from the original?
Each of the 120 strings that are 1 mutation away can mutate one OTHER nucleotide. Consider this one:
CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

It can't change the first position (that's already been found in another string in G1 (C → T and C→G) or G0 (C→A).

So there are 39 remaining positions to change. Each of them can be reached 2 ways. For example the string
CCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Can be reached by mutating position 1 in G1, and then position 2 in G2, or the reverse. Therefore, we have to divide our count by 2, because every second mutation will have been discovered 2 ways.

How do we calculate this?
Hint: It's NOT 120*120/2

Reminder here are are strings from G1
CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
GAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
TAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

ACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
ATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

How many strings are 2 mutations away from the original? Each of the 120 strings that are 1 mutation away can mutate one OTHER nucleotide. Consider this one:
CAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

It can't change the first position (that's already been found in another string in G1 (C → T and C→G) or G0 (C→A).

So there are 39 remaining positions to change. Each of them can be reached 2 ways. For example the string
CCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

Can be reached by mutating position 1 in G1, and then position 2 in G2, or the reverse. Therefore, we have to divide our count by 2, because every second mutation will have been discovered 2 ways.

How do we calculate this?
Hint: It's NOT 120*120/2

**Instead it's 40(3) * 39 (3) / 2 = 7020. Wagner asks how many have up to 2 mutations, so that's 7141 (7020+120+1 = 7141) The answer is NOT $7.2 \times 10^3$ as the paper reports.**

More generally, this is represented using choose notation. If you choose *k* out of *n* locations to mutate, and there are 3 different mutations for each location then, the number of possible genomes are

$\binom{n}{k} * 3^k$ where $\binom{n}{k} = \frac{n!}{k!(n-k)!}$

Ex: For n = 40, k = 2 (choose 2 out of 40 positions to mutate, each with 3 possible mutations):

$$\frac{40! * 3^k}{2!\,(40-2)!} = \frac{(40*39)*3^2}{2} = 7020$$

Ex: For n = 40, k = 1 (choose 1 out of 40 positions to mutate, each with 3 possible mutations):

$$\frac{40! * 3^k}{1!\,(40-1)!} = \frac{(40)*3}{1} = 120$$

Total number of genomes = 4^40.
Check that summing G0 to G40 gives you 4^40.

| k | =FACT(40)*3^k / (FACT(40-k)*FACT(k)) |
|---|---|
| 0 | 1 |
| 1 | 120 |
| 2 | 7020 |
| 3 | 266760 |
| 4 | 7402590 |
| 5 | 159895944 |
| 6 | 2798179020 |
| 7 | 40773465720 |
| 8 | 5.04572E+11 |
| 9 | 5.3821E+12 |
| 10 | 5.00535E+13 |
| 11 | 4.09529E+14 |
| 12 | 2.96908E+15 |
| 13 | 1.91848E+16 |
| 14 | 1.10998E+17 |
| 15 | 5.7719E+17 |
| 16 | 2.70558E+18 |
| 17 | 1.14589E+19 |
| 18 | 4.39258E+19 |
| 19 | 1.52584E+20 |

| k | =FACT(40)*3^k / (FACT(40-k)*FACT(k)) |
|---|---|
| 20 | 4.80641E+20 |
| 21 | 1.37326E+21 |
| 22 | 3.55799E+21 |
| 23 | 8.35355E+21 |
| 24 | 1.77513E+22 |
| 25 | 3.40825E+22 |
| 26 | 5.89889E+22 |
| 27 | 9.17605E+22 |
| 28 | 1.27809E+23 |
| 29 | 1.5866E+23 |
| 30 | 1.74526E+23 |
| 31 | 1.68896E+23 |
| 32 | 1.42506E+23 |
| 33 | 1.03641E+23 |
| 34 | 6.40134E+22 |
| 35 | 3.29212E+22 |
| 36 | 1.37171E+22 |
| 37 | 4.4488E+21 |
| 38 | 1.05366E+21 |
| 39 | 1.62102E+20 |
| 40 | 1.21577E+19 |

This is the 41 row Excel sheet to check that it works.

Phew, it works!

A genome of length 40 can have at most mutations in 40 locations.

All possible combinations of the 4 nucleotides is 4^40 = 1.20893E+24

Check: summing all possible genomes with 0 to 40 mutations gives this same answer.

| | |
|---|---|
| sum | 1.20893E+24 |

# The Adjacent Possible

[The Adjacent Possible Simplified]

*Biological proteins use 20 kinds of amino acids — glycine, alanine, lysine, arginine, and so forth. A protein is a linear sequence of these [, a string of amino acid beads.] Picture 20 colors of beads. A protein of 100 amino acids is like a string of 100 beads. The number of possible strings is just the number of types of beads, here 20, multiplied times itself 100 times. That's $10^{120}$ .. . equal to the square of the number of hydrogen molecules in the universe.*

The adjacent possible is a kind of shadow future, hovering on the edges of the present state of things, a map of all the ways in which the present can reinvent itself…

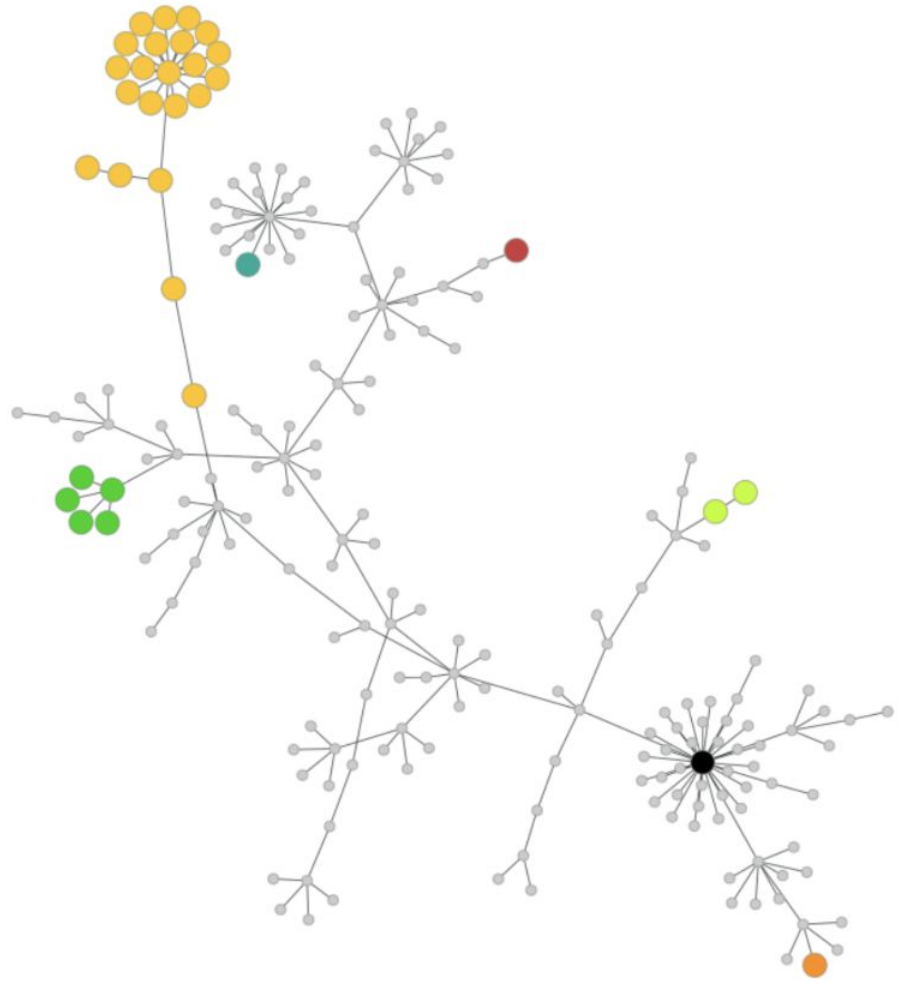


**Figure 2: A neutral network of 201 variants of the `look` utility. Each node in this graph represents a mutated variant of `look`. The original program is shown in black. All of the variants depicted pass all of the positive test cases (they retain all originally-tested functionality). Gray nodes are variants which exhibit the exact same behavior on the test suite as the original program. Some variants additionally pass a negative test case which fails for the original program (i.e., they repair a defect)—these are represented by colored nodes. Nodes in the graph which share a color fix the defect in the same way.**

# Epistasis in program repair?

## Epistasis in Software

- Deletion (595) is individually deleterious
- Insertion (718) is individually neutral
- Deletion (595) exposes Insertion (718)
- Combination repairs buffer overflow

```
594     bot = 0L;
595 -- fseek(dfile, 0L, 2);
596     top = ftell(dfile);
597     while (1) { ... }
...
...
...
710     else {
...
717     tmp___0 = tmp;
718 ++ return (tmp___1);
719     }
```

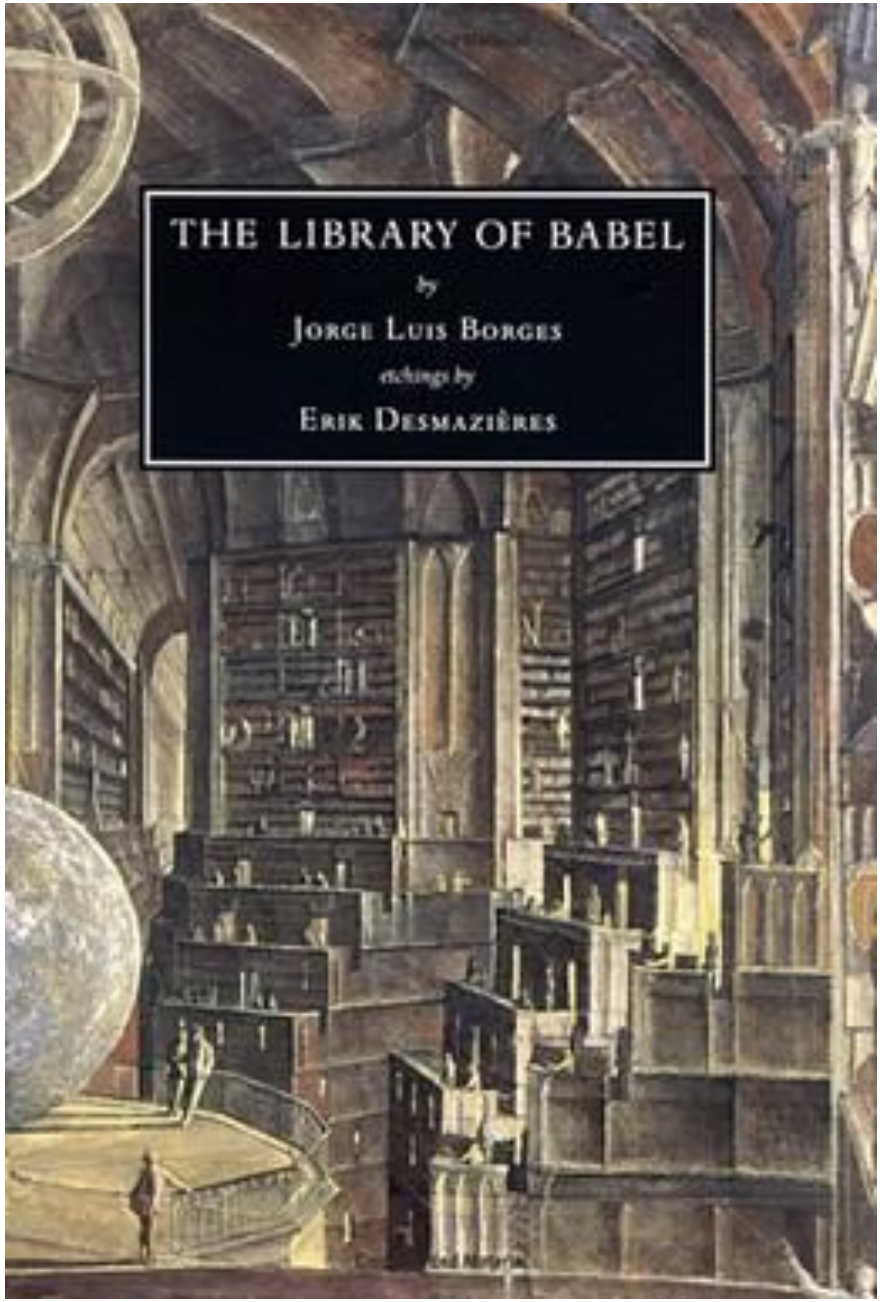
Code excerpt from `Ultrix look`



**Listing 1: Example Repair**

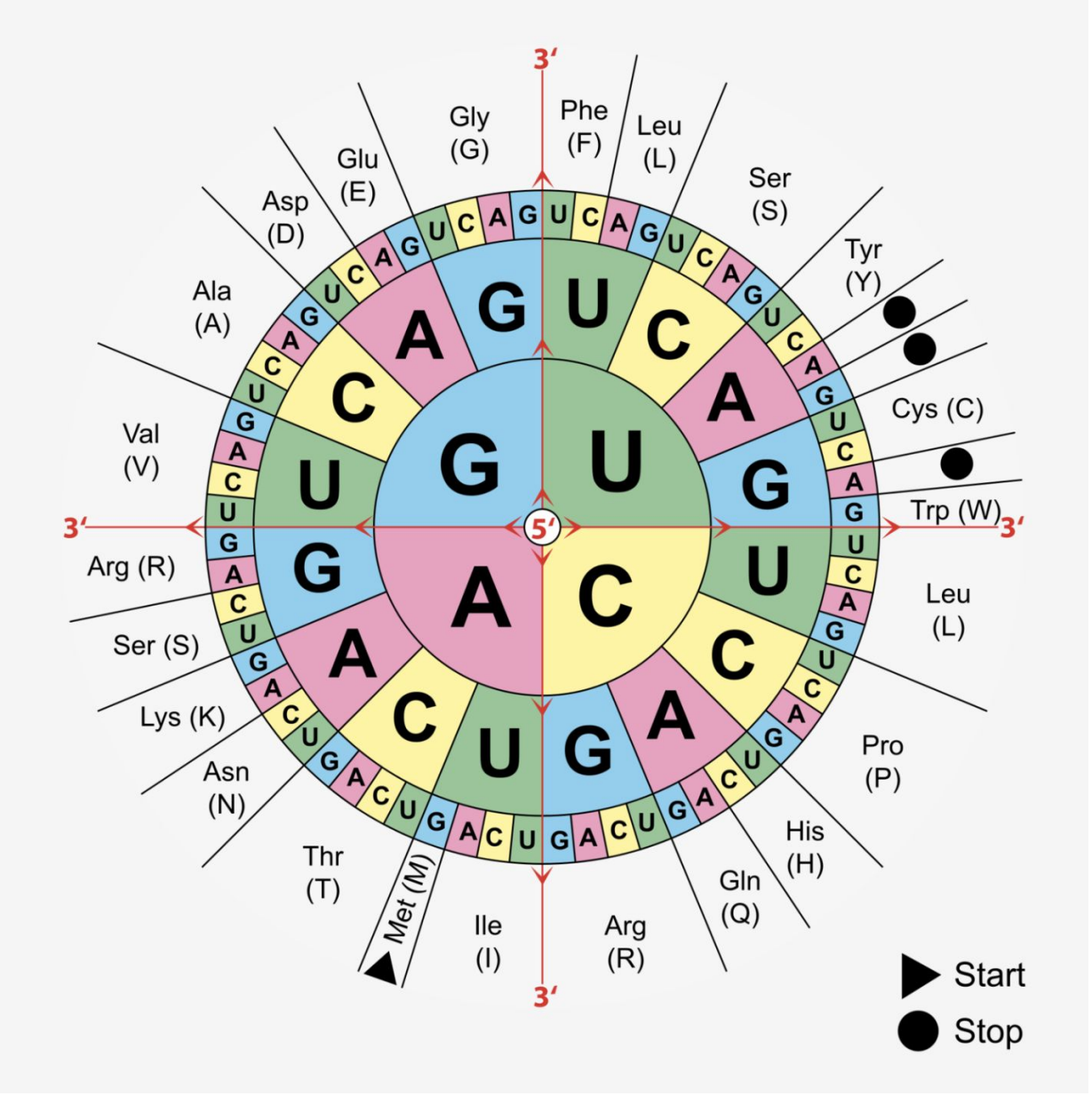
```
594     bot = 0L;
595 -- fseek(dfile, 0L, 2);
596     top = ftell(dfile);
597     while (1) { ... }
...
...
...
710     else {
...
717     tmp___0 = tmp;
718 ++ return (tmp___1);
719     }
```

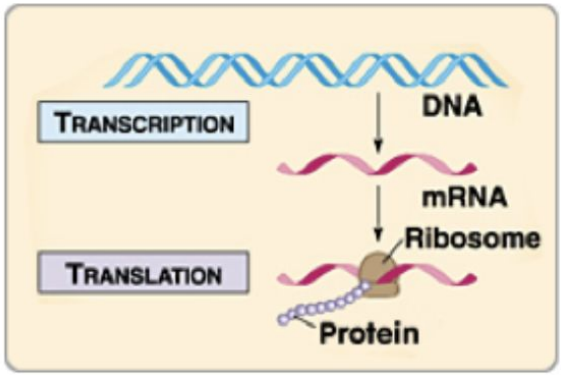
**The code excerpt above contains two edits to `look` that interact. When the program is fed a malformed input, the edit to line 595 changes the control flow of the program, causing the program to execute the return added by the edit to line 718 and exit. Without both of these edits, the program instead experiences a segmentation fault on the buggy input. The pair of edits repairs the bug.**




## Plot[edit]

Borges' narrator describes how his universe consists of an enormous expanse of adjacent hexagonal rooms. In each room, there is an entrance on one wall, the bare necessities for human survival on another wall, and four walls of bookshelves. Though the order and content of the books are random and apparently completely meaningless, the inhabitants believe that the books contain every possible ordering of just 25 basic characters (22 letters, the period, the comma, and space). Though the vast majority of the books in this universe are pure gibberish, the library also must contain, somewhere, every coherent book ever written, or that might ever be written, and every possible permutation or slightly erroneous version of every one of those books. The narrator notes that the library must contain all useful information, including predictions of the future, biographies of any person, and translations of every book in all languages. Conversely, for many of the texts, some language could be devised that would make it readable with any of a vast number of different contents.





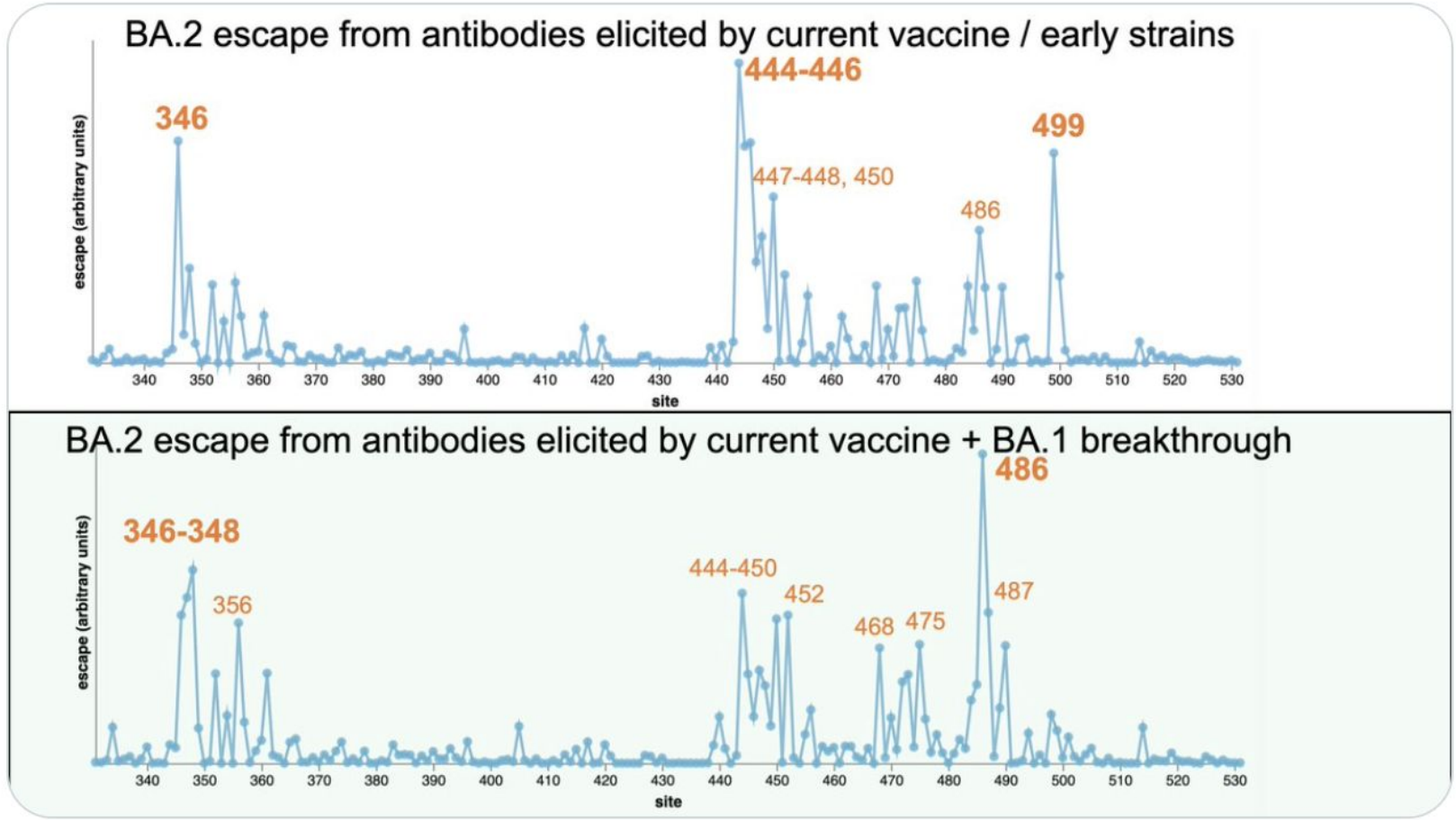

| | | | |
|---|---|---|---|
| A | Alanine | M | Methionine |
| C | Cisteine | N | Asparagine |
| D | Aspartic Acid | P | Proline |
| E | Glutamic Acid | Q | Glutamine |
| F | Phenylalanine | R | Arginine |
| G | Glycine | S | Serine |
| H | Histidine | T | Threonine |
| I | Isoleucine | V | Valine |
| K | Lysine | W | Tryptophan |
| L | Leucine | Y | Tyrosine |

Data from International Union of Pure and Applied Chemistry and International Union of Biochemistry.[70]

http://thebiologyprimer.com/chapter-transcription-translation

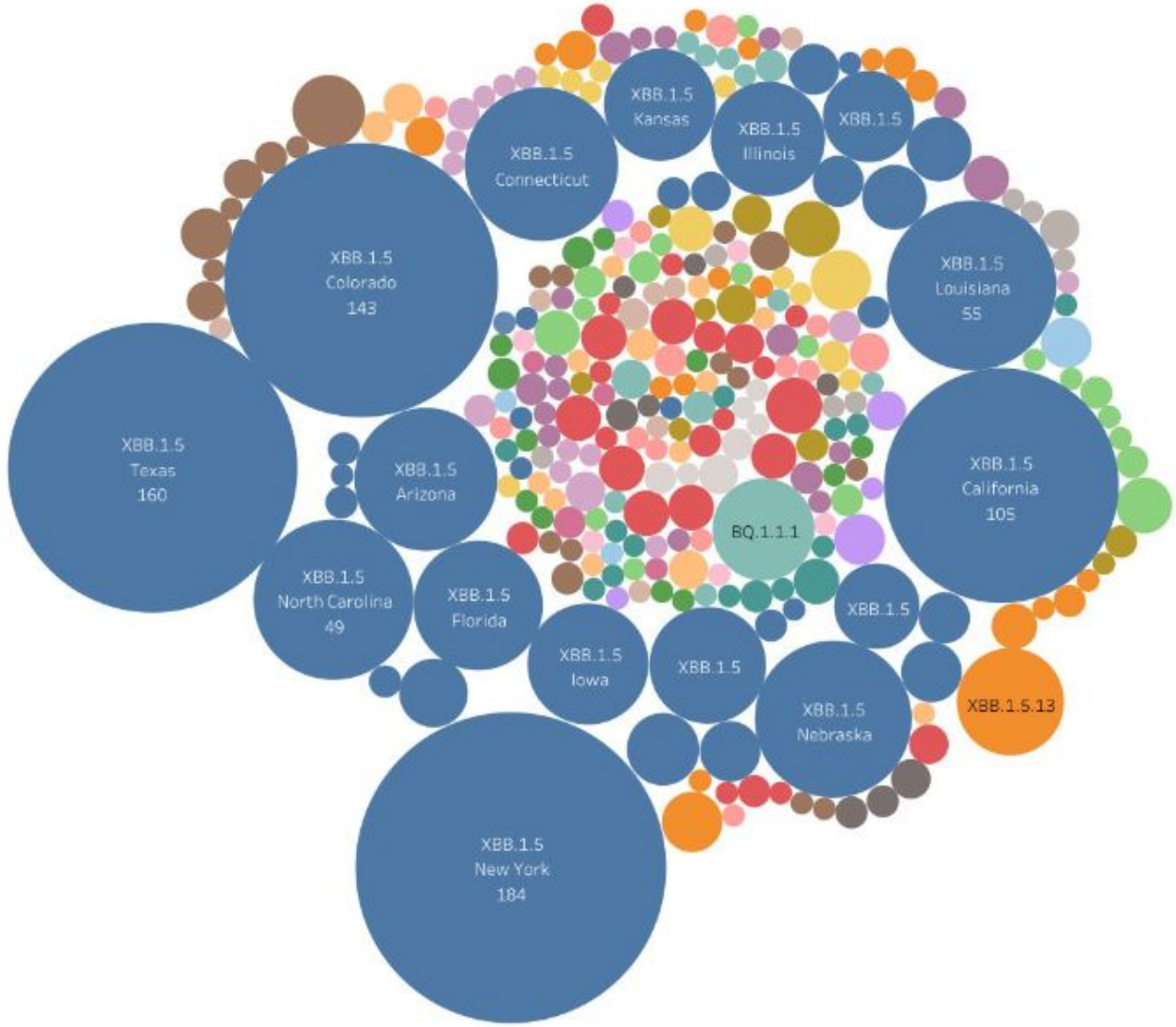
**Bloom Lab** @jbloom_lab · Jul 25, 2022

Here are new escape maps for BA.2. The major sites of potential future escape in BA.2 are 346-348, 444-446, 486, 499, and some more minor ones shown below. Note escape mutations differ a bit +/- a BA.1 breakthrough.



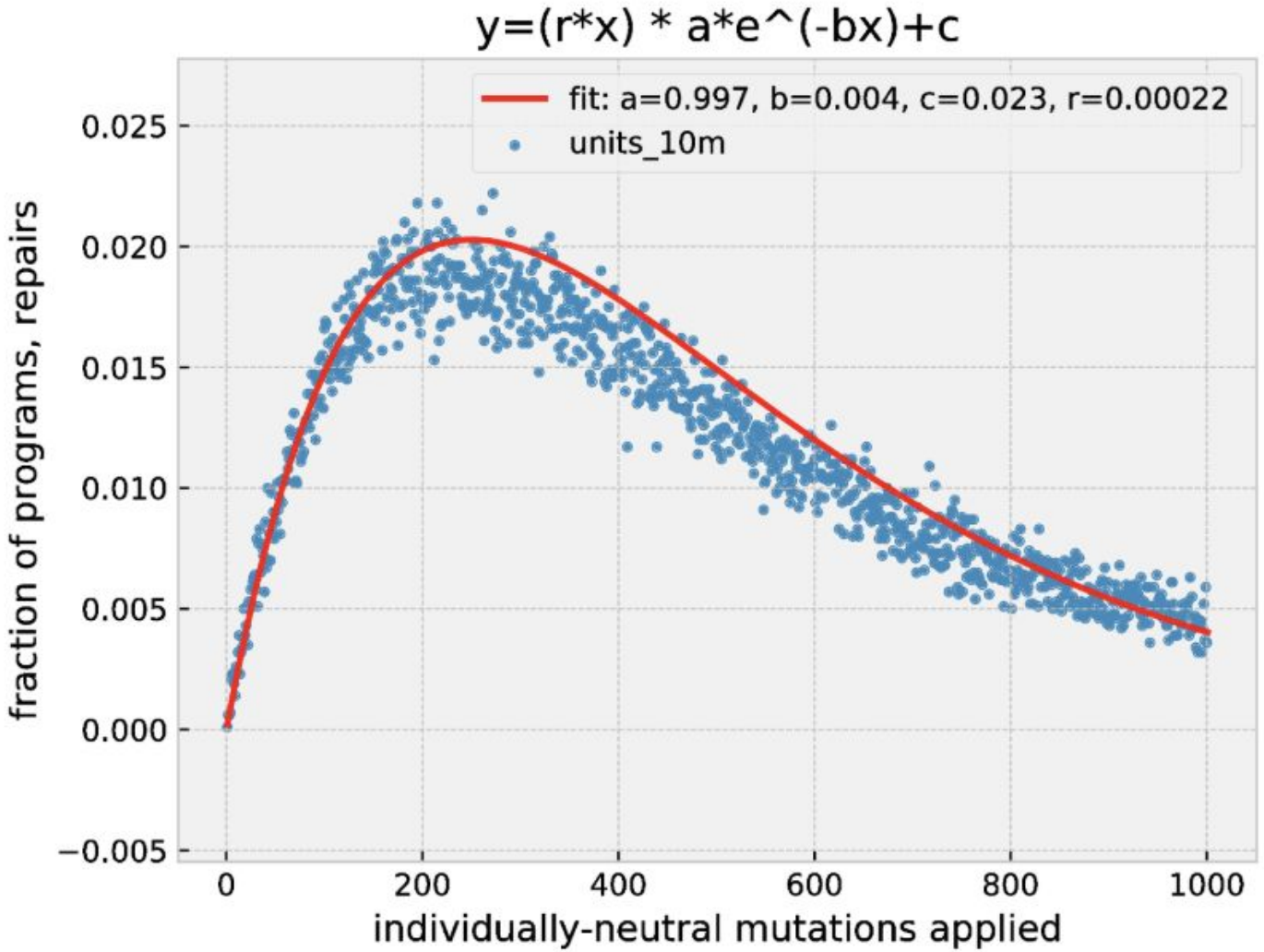

1 9 35

# SARSCoV2 Variant Dashboard - USA | 15-DAY TRENDS | NYITCOMResearch Report




| Pangolin Line.. | % of Total |
|---|---|
| XBB.1.5 | 69.67% |
| XBB.1.5.13 | 3.33% |
| BQ.1.1 | 2.75% |
| XBB.1.9.1 | 1.70% |
| BQ.1.1.1 | 1.37% |
| XBB.1.5.11 | 0.92% |
| BQ.1.22 | 0.85% |
| CH.1.1 | 0.78% |
| XBB.1 | 0.78% |
| BQ.1 | 0.72% |
| XBB | 0.65% |
| BQ.1.1.18 | 0.59% |
| BQ.1.1.32 | 0.59% |
| CH.1.1.1 | 0.52% |
| XBB.1.5.4 | 0.52% |
| XBB.1.5.5 | 0.52% |
| BQ.1.1.4 | 0.46% |
| XBB.1.15 | 0.46% |
| XBB.1.5.1 | 0.46% |
| BQ.1.1.5 | 0.39% |
| BQ.1.1.65 | 0.39% |
| XBB.1.5.7 | 0.39% |
| XBB.1.9.2 | 0.39% |
| XBB.2.3 | 0.39% |
| XBB.1.5.10 | 0.33% |
| XBB.1.5.2 | 0.33% |
| BQ.1.18 | 0.26% |
| BQ.1.25.1 | 0.26% |
| CK.1 | 0.26% |
| XBB.2.4 | 0.26% |
| BA.5.3.1 | 0.20% |
| BE.9 | 0.20% |
| BF.7 | 0.20% |
| BQ.1.1.10 | 0.20% |
| BQ.1.1.2 | 0.20% |
| BQ.1.1.40 | 0.20% |
| BQ.1.1.41 | 0.20% |
| BQ.1.1.63 | 0.20% |
| BQ.1.1.68 | 0.20% |
| BQ.1.1.69 | 0.20% |
| BQ.1.1.7 | 0.20% |
| XBB.1.5.12 | 0.20% |
| XBB.1.5.3 | 0.20% |
| XBB.1.5.6 | 0.20% |
| XBB.1.9.3 | 0.20% |
| BA.2 | 0.13% |
| BA.5.2.1 | 0.13% |

| Circulating Variants in the following US States :All - Specimen Collected in the last 15 days | **Updated on 3/2/2023 2:17:38 AM** | **Source (sequences): GISAID - ** Data is dynamic and is constantly being updated | Lineage assignments: Nextclade tool^^** | Summary of PANGO designations: https://www.pango.network/summary-of-designated-omicron-lineages/ |




J. Renzullo, work in progress