Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Research Article

Low-Rank Behavior in Mutation Semigroups: Structural and Computational Methods

M. I. Sampson[1], Reny George[2*], R. B. Abubakar[3], Julie S. George[4]

[1]Department of Mathematics, Akwa Ibom State University, Ikot Akpaden, 524101, Nigeria
[2]Department of Mathematics, College of Science and Humanities in Alkharj, Prince Sattam bin Abdulaziz University, Alkharj, 11942, Saudi Arabia
[3]Department of Mathematics and Statistics, Federal University Otuoke, Otuoke, Bayelsa State, 562101, Nigeria
[4]Department of Health and Rehabilitation Sciences, College of Applied Medical Sciences, Prince Sattam bin Abdulaziz University, Alkharj, 11942, Saudi Arabia
E-mail: r.kunnelchacko@psau.edu.sa

Received: 6 November 2025; **Revised:** 15 December 2025; **Accepted:** 31 December 2025

Abstract: We present an algebraic framework for analyzing mutation processes by encoding elementary operations as total maps on a finite set and organizing them into a semigroup. Central to our approach is the study of low-rank behavior: we characterize the existence of constant and small-rank elements in the mutation semigroup, linking them to synchronizing words in the associated deterministic automaton. Structural insights are obtained via Green's relations, image and kernel characterizations, and rank considerations. On the computational side, we employ the pair-graph method to test efficiently for constant maps and a breadth-first search on the power-set automaton to detect elements of bounded rank. A worked example on a three-element set illustrates these techniques and demonstrates the explicit discovery of rank-two and rank-one maps. Practical heuristics, such as the contraction of image sets, are also discussed to enhance efficiency in larger systems. The integration of algebraic, structural, and computational perspectives thus provides a coherent toolkit for identifying and understanding low-rank phenomena in mutation semigroups, and lays the groundwork for further characterization of generators that enforce such behavior. Finally, we extend the finite theory to parameterized and infinite families of mutations, drawing connections with quasispecies models in biology and interpreting image contractions as mechanisms of error suppression and genomic stability. By combining algebraic definitions, structural theorems, and algorithmic analyses, we provide a refined toolkit for understanding mutation collapse and its theoretical and biomedical implications.

***Keywords*:** mutation semigroups, rank collapse, k-transitive subgroups, algebraic synchronization, semigroup generators, epidemiological applications, genomic evolution

MSC: 20M10, 68Q45, 68Q70, 68W30, 92C50, 92D20

1. Introduction

This paper aims to supply an algebraic account of mutational processes in genomics, emphasizing finite semigroup structure rather than the popularly used operator-analytic or probabilistic models [1, 2]. The principal object of study is the

Copyright ©2026 Reny George, et al.
DOI: https://doi.org/10.37256/cm.7220269094
This is an open-access article distributed under a CC BY license
(Creative Commons Attribution 4.0 International License)
https://creativecommons.org/licenses/by/4.0/

mutation semigroup generated by a finite set of elementary mutation maps acting on a finite collection of genome states. In this context, a *mutation* is understood as a transformation $\mu: X \to X$ on the finite state space X of genome configurations, representing a heritable alteration in one or more loci or parameters of the system. Each elementary mutation acts as a deterministic map (or, more generally, a stochastic operator) modifying genomic states according to specified rules of substitution, insertion, or deletion. The collection of all such transformations, closed under composition, constitutes the mutation semigroup.

The significance of this study lies in providing a rigorous algebraic framework that can bridge the gap between abstract semigroup theory and practical genomic modeling, offering a new perspective on mutational dynamics, resilience, and evolutionary bottlenecks. By formalizing mutation processes in terms of semigroups, the work enables precise structural analysis, algorithmic detection of invariants, and potential predictive modeling that is both mathematically robust and biologically interpretable.

From a theoretical perspective, this work introduces the class of *mutation semigroups*—finite semigroups generated by biologically constrained transformations—which exhibit new structural properties concerning rank, minimal ideals, and synchronization. The classification of generating sets under such constraints and the derivation of transitivity-based synchronization criteria represent previously unstudied phenomena in semigroup theory. These results are entirely theoretical and contribute to the algebraic understanding of semigroups beyond their biological interpretation.

Our main objectives are to provide clear and rigorous algebraic definitions accessible to both algebraists and mathematically inclined biologists, to establish structural theorems that connect semigroup-theoretic invariants with biologically relevant notions such as resistance as an absorbing phenomenon [3], and to develop examples and algorithms that enable practical implementation and testing of the theory on real or simulated datasets.

To achieve these objectives, the paper is structured as follows. Section 2 lays out the algebraic foundations of mutation semigroups, presenting precise definitions, lemmas, and structural theorems that link rank, invariance, and collapse. Section 3.1 describes algorithmic approaches, including the pair-graph collapse test and invariant-partition detection, along with complexity analyses and heuristic strategies inspired by synchronization-control methods in cellular automata [4]. Section 3.2 presents classification results for special families of generators, culminating in transitivity-based criteria—ranging from full symmetric groups to k-transitive subgroups—that guarantee synchronization. Section 3.4 then interprets these algebraic results in a biological context, drawing analogies between rank reduction and mutational bottlenecks, and introducing the concept of resilience thresholds in genomic evolution.

Section 3.4 connects ideals and minimal ideals with biologically relevant absorbing states. Section 3.5 formalizes evolutionary equivalence via congruences, while Section 3.6 applies Rees matrix representations to describe minimal-ideal structure. Section 3.9 introduces the problem of constant and low-rank maps, providing synchronization criteria and algorithmic tests. Section 3.10 characterizes generating sets enforcing synchronizing behavior, linking mutation semigroups to Černý's conjecture. Section 3.11 investigates enumeration and complexity, analyzing the rank spectrum of mutation families. Section 3.12 develops heuristic refinements, including contraction operators and probabilistic methods. Section 3.13 explores broader medical, synthetic, and epidemiological applications, and Section 4 concludes by highlighting theoretical open problems and biomedical implications. Together, these sections establish a comprehensive algebraic and computational framework for mutation semigroups, integrating semigroup theory, automata methods, and genomic interpretation.

All definitions and results that are standard in semigroup theory are as cited; proofs are given for the convenience of readers and to ensure self-contained rigor.

2. Preliminaries: semigroups and transformation semigroups

We introduce standard algebraic definitions used throughout the paper. Each definition is stated formally and compactly.

Notation 1 Throughout this paper, S denotes a semigroup. When it is necessary to ensure the existence of an identity element (for example, when defining principal ideals or morphisms that require unital extensions), we adjoin an identity 1 to S, forming the semigroup

$$S^1 = S \cup \{1\}, \quad \text{with } 1a = a1 = a \text{ for all } a \in S.$$

If S already has an identity, we take $S^1 = S$. All products and factorizations are taken with respect to the associative operation of the underlying semigroup.

Definition 1 (Genome) Let A be a finite alphabet (e.g. $\{0, 1\}$ or nucleotide sets [3, Ch. 1, 5, Ch. 2]). A *genome* of length n is an element of the state space

$$X = A^n.$$

Each $x = (a_1, \ldots, a_n) \in X$ represents one configuration of genetic sites.

Definition 2 (Mutation) A *mutation* on a genome state space X is a transformation

$$\mu \colon X \to X,$$

representing an allowed elementary operation (substitution, indel, duplication, transposition, etc. [6–9]). A finite family $M \subseteq \mathscr{T}_X$ generates the *mutation semigroup*

$$S = \langle M \rangle \subseteq \mathscr{T}_X,$$

the set of all finite compositions of admissible mutations; see [10, Ch. 1–3, 11, Vol. I].

Example 1 (Bit-flip mutation semigroup) Let $A = \{0, 1\}$ and $X = A^n$. For each i, the bit-flip map $\mu_i \colon X \to X$ defined by

$$(\mu_i(x))_j = \begin{cases} 1 - x_i, & j = i, \\ x_j, & j \neq i \end{cases}$$

models a simple substitution event [6, 12]. The semigroup $S = \langle \mu_1, \ldots, \mu_n \rangle$ consists of all finite sequences of coordinatewise flips.

Remark 1 The mutation semigroup S formalizes cumulative mutation pathways: composition corresponds to successive changes, while an identity (when adjoined) models a neutral step. These transformation semigroups form the basis for rank, kernel, and synchronization analyses developed later; see [10, Ch. 5, 13–15].

Definition 3 (Semigroup) A *semigroup* is a set S equipped with an associative operation

$$\cdot \colon S \times S \to S,$$

such that $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ for all $a, b, c \in S$. Foundational references include [10, 11, 16].

Definition 4 (Monoid) A *monoid* is a semigroup $(M, \cdot)$ that possesses an identity element $1 \in M$ such that $1 \cdot m = m = m \cdot 1$ for all $m \in M$ [11].

Definition 5 (Idempotent and unit) Let $(S, \cdot)$ be a semigroup. An element $e \in S$ is *idempotent* if $e^2 = e$. An element $u \in S$ is a *unit* if there exists $v \in S$ such that $uv = vu = 1$ (assuming S is a monoid); the set of all units is denoted $U(S)$ [11].

Definition 6 (Transformation semigroup) Let X be a nonempty set. The *full transformation semigroup* on X, denoted $\mathscr{T}_X$, is the set of all functions $f: X \to X$ equipped with composition of functions as the semigroup operation (compose right-to-left). A *transformation semigroup* on X is any subsemigroup of $\mathscr{T}_X$ [10, 17].

Definition 7 (Free semigroup and presentation [10, 11]) Let M be a nonempty set (the *alphabet*). The *free semigroup* on M, denoted M^+, is the set of all nonempty finite words over M with concatenation as the operation. A *semigroup presentation* is an ordered pair $\langle M \mid R \rangle$ where $R \subseteq M^+ \times M^+$ is a set of defining relations. The semigroup defined by the presentation is the quotient $M^+/\sim_R$, where $\sim_R$ is the smallest congruence on M^+ containing R.

Example 2 (Free semigroup) Let $M = \{a, b\}$. Then

$$M^+ = \{a, b, aa, ab, ba, bb, aaa, aab, \ldots\}$$

is the set of all nonempty finite strings (words) over $\{a, b\}$, with concatenation as the operation; for instance $ab \cdot ba = abba$. Every distinct word represents a distinct element—no simplifications or identifications are made. In biological terms, a and b could represent two distinct types of elementary mutations or editing operations, and each word corresponds to a possible sequence of such mutations applied in order.

Example 3 (Semigroup presentation) [18] Let $M = \{a, b\}$ and let $R = \{(ab, ba)\}$. This relation identifies the two words ab and ba, meaning that the order of a and b no longer matters—they commute. Hence, in the resulting quotient $M^+/\sim_R$, any word is characterized only by the number of a's and b's it contains, not by their arrangement. Algebraically, this semigroup is the *free commutative semigroup* on two generators, which can be represented as

$$S = \{a^i b^j : i, j \in \mathbb{N},\ (i, j) \neq (0, 0)\},$$

with multiplication $(i_1, j_1) \cdot (i_2, j_2) = (i_1 + i_2, j_1 + j_2)$. Intuitively, this means we are counting how many times each elementary mutation occurs, rather than tracking their exact order of application—a natural abstraction when mutations act independently on separate genetic sites.

Remark 2 Free semigroups capture all possible ordered compositions of generators, while presentations impose specific algebraic constraints by identifying equivalent words. In modeling mutation processes, these concepts provide a rigorous way to describe how different mutation operators combine—either freely (distinct orders matter) or under simplifying biological assumptions such as commutativity (order does not matter).

Definition 8 (Congruence ([11])) Let $(S, \cdot)$ be a semigroup. An equivalence relation $\equiv$ on S is a *congruence* if for all $a, b, c, d \in S$ with $a \equiv b$ and $c \equiv d$ we have $ac \equiv bd$. The quotient $S/\equiv$ is naturally a semigroup (where the symbol "$\equiv$" means we are identifying elements of S that are equivalent under a congruence relation $\equiv$.

Remark 3 The quotient set

$$S/\equiv = \{[a] : a \in S\}$$

is the set of all equivalence classes of elements of S under $\equiv$.

Each class $[a]$ is defined as

$$[a] = \{b \in S : b \equiv a\}.$$

The operation on $S/\equiv$ is induced by the operation on S. It is defined by

$$[a] \cdot [b] = [ab].$$

This is well-defined because $\equiv$ is a congruence—it guarantees that the product does not depend on the choice of representatives from the classes.

Definition 9 (Left, right and two-sided ideals) Let S be a semigroup. A nonempty subset $I \subseteq S$ is a *left ideal* if $SI \subseteq I$. It is a *right ideal* if $IS \subseteq I$. It is a (two-sided) *ideal* if it is both a left and right ideal [11].

Definition 10 (Minimal ideal) A nonempty ideal K of S is *minimal* if K does not properly contain any nonempty ideal of S. The minimal ideal of a finite semigroup exists and is unique; we denote it by $K(S)$ [10].

The classical Green's relations $\mathscr{L}, \mathscr{R}, \mathscr{J}, \mathscr{H}, \mathscr{D}$ are standard tools to study semigroup structure [10, 11].

Definition 11 (Green's relations) Let S be a semigroup and let S^1 denote S with an identity adjoined if necessary (see notation 1). Define binary relations $\mathscr{L}, \mathscr{R}, \mathscr{J}$ on S by:

$$a \mathscr{L} b \iff S^1 a = S^1 b,$$

$$a \mathscr{R} b \iff a S^1 = b S^1,$$

$$a \mathscr{J} b \iff S^1 a S^1 = S^1 b S^1.$$

The relations $\mathscr{H}$ and $\mathscr{D}$ are defined by $\mathscr{H} = \mathscr{L} \cap \mathscr{R}$ and $\mathscr{D} = \mathscr{L} \circ \mathscr{R} = \mathscr{R} \circ \mathscr{L}$ (their composition as relations).

Example 4 Let $S = \{a, b, c\}$ with multiplication table

$\cdot$	a	b	c
a	a	b	c
b	b	b	c
c	c	c	c

Then:

• $aS = \{a, b, c\}$, $bS = \{b, c\}$, $cS = \{c\}$. Hence $\mathscr{R}$-classes are $\{a\}$, $\{b\}$, $\{c\}$ (each element generates a distinct right ideal).

• $Sa = \{a, b, c\}$, $Sb = \{b, c\}$, $Sc = \{c\}$. Thus the $\mathscr{L}$-classes coincide with the $\mathscr{R}$-classes.

Therefore,

$$\mathscr{H} = \mathscr{L} \cap \mathscr{R}$$

has the same partition:

$$\mathscr{H} = \{\{a\}, \{b\}, \{c\}\}.$$

In this example, each $\mathscr{H}$-class is a singleton, reflecting that no two distinct elements generate the same left and right ideals simultaneously. In more structured semigroups, nontrivial $\mathscr{H}$-classes form maximal subgroups.

Example 5 Let $S = \{a, b, c, d\}$ with the multiplication table

$\cdot$	a	b	c	d
a	a	b	c	d
b	b	b	c	d
c	c	c	c	d
d	d	d	d	d

The $\mathscr{L}$-classes are $\{a, b\}$, $\{c\}$, $\{d\}$, and the $\mathscr{R}$-classes are $\{a\}$, $\{b, c\}$, $\{d\}$. Then

$$\mathscr{D} = \mathscr{L} \circ \mathscr{R} = \mathscr{R} \circ \mathscr{L} = \{\{a, b, c\}, \{d\}\}.$$

Thus $\mathscr{D}$ merges $\mathscr{L}$- and $\mathscr{R}$-classes that are connected through either relation, producing the coarsest partition compatible with both.

Remark 4 In finite semigroups $\mathscr{D} = \mathscr{J}$; this will be used below when S is finite.

3. Main results

3.1 *A formal model: finite genome states and elementary mutation maps*

We provide a mathematically precise model—the *mutation semigroup* $S = \langle G \rangle \subseteq \mathscr{T}_X$, generated by a finite set of mutation maps acting on the genome state space X.

Definition 12 (Alphabet and genomes) Let Σ be a finite alphabet of size $q \geq 2$. Fix $N \in \mathbb{N}$, $N \geq 1$. The set of *genomes* under consideration is

$$X := \Sigma^{\leq N} = \bigcup_{\ell=0}^{N} \Sigma^{\ell},$$

the set of all words over Σ of length at most N. (We include the empty word ε as length 0.) Note that X is finite.

Example 6 Let $\Sigma = \{A, T, G, C\}$ and $N = 2$. Then

$$X = \Sigma^{\leq 2} = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 = \{\varepsilon\} \cup \{A, T, G, C\} \cup \{AA, AT, \ldots, CC\},$$

where ε denotes the empty word. Thus, X is the set of all DNA fragments of length at most two.

We restrict attention to this finite set X so that all transformation semigroups we construct are finite; this simplifies structural algebraic results while remaining biologically meaningful for studies limited to genomes up to a maximum length N.

Remark 5 (Elementary operations) Before introducing mutation maps formally, we recall the kinds of primitive operations we allow on strings. These include:

(i) substitution at a fixed position [6, 12],

(ii) deletion at a fixed position (when defined) [7, 12],

(iii) insertion of a specified letter at a specified position (with truncation rule) [7, 12],

(iv) duplication of a substring [8, 14], and

(v) transposition of adjacent letters [9, 12, 15, 19].

Each such operation is intended to act on the entire state space X in a well-defined manner.

We denote by $\mathscr{M}$ a finite collection of such operations, each of which we formalize as an elementary mutation map.

Definition 13 (Total function) Let X and Y be nonempty sets. A function $f: X \to Y$ is called *total* if it assigns exactly one value $f(x) \in Y$ to every element $x \in X$. That is, f is defined on all of X.

Definition 14 (Elementary mutation maps) An *elementary mutation map* is a total function $\mu: X \to X$ such that μ corresponds to one of the elementary operations described above. Thus every $\mu \in \mathscr{M}$ is a well-defined endomap on X, ensuring closure under application.

Definition 15 (Mutation action and the mutation semigroup) Let the set $\mathscr{M}$ acts on X by evaluation: for $\mu \in \mathscr{M}$ and $x \in X$ we write $\mu(x) \in X$ for the image. Let the action be extended to words over $\mathscr{M}$ by composing maps: if $w = \mu_1 \mu_2 \cdots \mu_k \in \mathscr{M}^+$ then define

$$w(x) := (\mu_1 \circ \mu_2 \circ \cdots \circ \mu_k)(x).$$

The *mutation semigroup* generated by $\mathscr{M}$ is the subsemigroup

$$S := \langle \mathscr{M} \rangle \leq \mathscr{T}_X$$

consisting of all finite compositions of elements of $\mathscr{M}$ (with the semigroup operation given by function composition).

Example 7 (A simple mutation semigroup) Let $\Sigma = \{A, T\}$ and $X = \Sigma^{\leq 3}$. Define three elementary mutation maps on X:

$$\mu_1: \text{replace } A \text{ by } T, \quad \mu_2: \text{delete the last letter (if any)}, \quad \mu_3: \text{insert } A \text{ at the end.}$$

Then the mutation semigroup

$$S = \langle \mathscr{M} \rangle = \langle \mu_1, \mu_2, \mu_3 \rangle \leq \mathscr{T}_X$$

consists of all finite compositions of these maps. For instance,

$$\mu_2 \circ \mu_1(ATA) = TT, \qquad \mu_3 \circ \mu_2(TA) = TAA.$$

Thus S forms a finite transformation semigroup on X under function composition, capturing the combined effects of substitution, deletion, and insertion operations.

Remark 6 Because X is finite and $\mathscr{M}$ is finite, S is a finite semigroup. In particular S has a unique minimal ideal $K(S)$.

3.2 *Presentations and the kernel congruence*

We formalize the relationship between the free semigroup on $\mathscr{M}$ and the concrete transformation semigroup S.

Definition 16 (Action homomorphism and kernel congruence) Define the action homomorphism

$$\varphi\colon \mathscr{M}^{+} \longrightarrow \mathscr{T}_X, \qquad \varphi(w) = \text{the transformation } (x \mapsto w(x)).$$

The *kernel congruence* of φ is

$$\ker(\varphi) := \{(u, v) \in \mathscr{M}^{+} \times \mathscr{M}^{+} \colon \varphi(u) = \varphi(v)\}.$$

Example 8 (Kernel congruence) Let $\mathscr{M} = \{\mu_1, \mu_2\}$, where μ_1 and μ_2 are mutation maps on X defined by

$$\mu_1\colon \text{replace } A \text{ by } T, \qquad \mu_2\colon \text{insert } A \text{ at the end.}$$

Define the homomorphism

$$\varphi\colon \mathscr{M}^{+} \to \mathscr{T}_X, \qquad \varphi(w) = \text{the transformation on } X \text{ obtained by composing the maps in } w.$$

Then $(\mu_1\mu_2, \mu_2\mu_1) \in \ker(\varphi)$ whenever

$$\mu_1\mu_2(x) = \mu_2\mu_1(x) \quad \text{for all } x \in X.$$

For example, if $X = \{A, T\}$, both compositions send $A \mapsto TA$ and $T \mapsto TT$, so they represent the same transformation. Hence, $\mu_1\mu_2 \equiv \mu_2\mu_1 \pmod{\ker(\varphi)}$, showing that the kernel congruence identifies different words in $\mathscr{M}^{+}$ that act identically on X.

Proposition 1 Let S be a semigroup. The kernel $\ker(\varphi)$ is a semigroup congruence (see 8) on $\mathscr{M}^{+}$, and there is an isomorphism of semigroups

$$\mathscr{M}^{+}/\ker(\varphi) \cong S.$$

Consequently S admits the presentation $\langle \mathscr{M} \mid R \rangle$ where R is any set of relations generating $\ker(\varphi)$.

Proof. The map φ is a semigroup homomorphism by construction because concatenation of words corresponds to composition of functions. The kernel of a homomorphism is always a congruence, so $\ker(\varphi)$ is a congruence. The First Isomorphism Theorem for semigroups yields

$$\mathscr{M}^{+}/\ker(\varphi) \cong \operatorname{im}(\varphi) = S,$$

which proves the proposition. The last statement about presentations follows from the definition of presentation through a generating congruence. □

3.3 *Green's relations in the full transformation semigroup and consequences for mutation semigroups*

We recall how Green's relations specialize in $\mathscr{T}_X$ when X is finite, and we prove the standard characterizations. These results will be applied to elements of $S \subseteq \mathscr{T}_X$.

Definition 17 (Image, kernel (partition), rank) Let $f\colon X \to X$ be a transformation. The *image* of f is $\mathrm{im}(f) := f(X)$. The *kernel* of f is the equivalence relation on X given by

$$x \sim_f y \iff f(x) = f(y).$$

The *rank* of f is $\mathrm{rank}(f) := |\mathrm{im}(f)|$ (the image of the function f).

Definition 18 (low-rank elements) Let X be a finite set and $f \in \mathscr{T}_X$ a transformation. An element of a mutation semigroup $S \leq \mathscr{T}_X$ is called *low-rank* if $\mathrm{rank}(f)$ is strictly smaller than $|X|$. In particular, constant maps correspond to rank one, and more generally rank-r maps with $r \ll |X|$ represent strong contraction phenomena within the mutation system.

Theorem 1 (Green's relations in $\mathscr{T}_X$ for finite X) Let X be a finite set and let $f, g \in \mathscr{T}_X$. Then:

(1) $f \mathscr{R} g$ if and only if $\mathrm{im}(f) = \mathrm{im}(g)$.

(2) $f \mathscr{L} g$ if and only if $\ker(f) = \ker(g)$.

(3) $f \mathscr{J} g$ if and only if $\mathrm{rank}(f) = \mathrm{rank}(g)$.

Proof. We prove each part.

(1) Suppose $\mathrm{im}(f) = \mathrm{im}(g)$. Define $t\colon X \to X$ as follows: for each $x \in X$ choose $t(x) \in X$ such that $f(t(x)) = g(x)$ (possible since $g(x) \in \mathrm{im}(g) = \mathrm{im}(f)$). Then for all $x \in X$ we have $(f \circ t)(x) = f(t(x)) = g(x)$, hence $f \circ t = g$, i.e. $g \in f\mathscr{T}_X^1$. By symmetry there exists s with $f = g \circ s$, so $f \in g\mathscr{T}_X^1$. Therefore $f\mathscr{T}_X^1 = g\mathscr{T}_X^1$, which means $f \mathscr{R} g$. Conversely, if $f \mathscr{R} g$ then $f = g \circ t$ and $g = f \circ s$ for some $s, t \in \mathscr{T}_X^1$. This implies $\mathrm{im}(f) \subseteq \mathrm{im}(g)$ and $\mathrm{im}(g) \subseteq \mathrm{im}(f)$, so the images coincide.

(2) Suppose $\ker(f) = \ker(g)$. Define $u\colon \mathrm{im}(f) \to X$ by choosing for each $y \in \mathrm{im}(f)$ an element $u(y) \in X$ such that $g(u(y)) = y$ if possible. A more direct construction is as follows. Define map $s\colon X \to X$ by the rule: for each $z \in X$ pick an element $s(z) \in X$ such that $f(s(z)) = g(z)$. The kernel equality ensures that the assignment $f(s(z))$ depends only on $f(z)$, and therefore the map s is well-defined and satisfies $f \circ s = g$. Thus $g \in \mathscr{T}_X^1 f$. By symmetry $f \in \mathscr{T}_X^1 g$ and therefore $\mathscr{T}_X^1 f = \mathscr{T}_X^1 g$, i.e. $f \mathscr{L} g$. The converse implication is similar: if $\mathscr{T}_X^1 f = \mathscr{T}_X^1 g$ then there are maps p, q with $g = p \circ f$ and $f = q \circ g$, and one checks that $x \sim_f y \iff x \sim_g y$.

(3) For finite X the relations $\mathscr{D}$ and $\mathscr{J}$ coincide. We show that $\mathscr{J}$-classes are determined by rank. If $\mathrm{rank}(f) = \mathrm{rank}(g) = r$ then, picking enumerations of the images, one can construct maps $u, v \in \mathscr{T}_X$ such that $g = u \circ f \circ v$ (the construction first restricts f to a transversal of its image and then maps onto the image of g by a bijection). Hence $g \in \mathscr{T}_X^1 f \mathscr{T}_X^1$ and symmetrically $f \in \mathscr{T}_X^1 g \mathscr{T}_X^1$, so $f \mathscr{J} g$. Conversely, if $f \mathscr{J} g$ then $\mathrm{rank}(f) = \mathrm{rank}(g)$ because composing with maps cannot change the minimal cardinality of images appearing in the two-sided ideal. □

Corollary 1 Let X be finite and let $S \leq \mathscr{T}_X$ be a subsemigroup of the full transformation semigroup. Denote Green's relations computed inside S by $\mathscr{R}_S$, $\mathscr{L}_S$, $\mathscr{J}_S$, $\mathscr{H}_S$. Then:

1. (Full-transformation case) If $S = \mathscr{T}_X$ then for all $a, b \in S$ we have

$$a \mathscr{R} b \iff \mathrm{im}(a) = \mathrm{im}(b),$$

$$a \mathscr{L} b \iff \ker(a) = \ker(b),$$

$$a \mathscr{J} b \iff \operatorname{rank}(a) = \operatorname{rank}(b),$$

and $\mathscr{H} = \mathscr{R} \cap \mathscr{L}$ accordingly.

2. (General subsemigroup) For a general subsemigroup $S \leq \mathscr{T}_X$ the image/kernel/rank equalities are necessary but not always sufficient for the corresponding Green relation inside S. More precisely, for $a, b \in S$:

- If $a \mathscr{R}_S b$ then $\operatorname{im}(a) = \operatorname{im}(b)$.
- If $a \mathscr{L}_S b$ then $\ker(a) = \ker(b)$.
- If $a \mathscr{J}_S b$ then $\operatorname{rank}(a) = \operatorname{rank}(b)$.

Conversely, each equality characterisation is sufficient for the corresponding Green relation in S exactly when S contains the appropriate "witness" transformations (right/left/bi-multipliers) that realize the necessary factorizations. Concretely, for the $\mathscr{R}_S$-relation:

$$\operatorname{im}(a) = \operatorname{im}(b) \quad \text{and} \quad \exists r, s \in S \text{ with } ar = b,\ bs = a \quad \Longrightarrow \quad a \mathscr{R}_S b.$$

Analogous statements hold for $\mathscr{L}_S$ (with left multipliers) and for $\mathscr{J}_S$ (with two-sided multipliers).

Proof. We split the proof into two parts.

(1) The full-transformation case $S = \mathscr{T}_X$. These are standard classical facts; we sketch the arguments.

$\mathscr{R}$-case. In $\mathscr{T}_X$ the principal right ideal of a is $a\mathscr{T}_X = \{a \circ t \colon t \in \mathscr{T}_X\}$. If $\operatorname{im}(a) = \operatorname{im}(b)$ then there is, for each $y \in \operatorname{im}(a)$, a choice of $x \in X$ with $a(x) = y$ and a choice x' with $b(x') = y$. Define $r \in \mathscr{T}_X$ arbitrarily on the complement of $\operatorname{im}(a)$ and on $\operatorname{im}(a)$ set $r(y) = x'$ (extend to a full map). Then $a \circ r = b$. Symmetrically one finds s with $b \circ s = a$. Hence $a\mathscr{T}_X = b\mathscr{T}_X$ and $a \mathscr{R} b$. Conversely, if $a \mathscr{R} b$ then $a\mathscr{T}_X = b\mathscr{T}_X$, so there is r with $a \circ r = b$; then $\operatorname{im}(b) \subseteq \operatorname{im}(a)$. Symmetry yields equality.

$\mathscr{L}$-case. Analogous arguments with left-multipliers yield the kernel characterization: in $\mathscr{T}_X$ two maps generate the same left ideal iff they have the same kernel (equivalence relation induced by fibres).

$\mathscr{J}$-case. In $\mathscr{T}_X$ the two-sided ideal generated by a depends only on $\operatorname{rank}(a) = |\operatorname{im}(a)|$; every map of a given rank can be written as $u \circ a \circ v$ for suitable $u, v \in \mathscr{T}_X$. Hence $\mathscr{J}$ coincides with equal rank.

Combining $\mathscr{R}$ and $\mathscr{L}$ gives the $\mathscr{H}$-characterisation.

(2) The general subsemigroup $S \leq \mathscr{T}_X$. We first prove the necessary directions (these do not require extra hypotheses).

Necessity. Suppose $a \mathscr{R}_S b$ in S. By definition $aS = bS$ (the principal right ideals computed inside S). In particular there exists $r \in S$ with $a \circ r = b$. Thus $\operatorname{im}(b) \subseteq \operatorname{im}(a)$. By symmetry $\operatorname{im}(a) \subseteq \operatorname{im}(b)$; hence $\operatorname{im}(a) = \operatorname{im}(b)$. The $\mathscr{L}_S$ and $\mathscr{J}_S$ necessary directions are identical (replace right- by left- or two-sided ideals and use left or two-sided multipliers). This proves the one-way implications.

Sufficiency (with witness multipliers). The converse need not hold in general because the right/left/two-sided multipliers that realize the factorizations in $\mathscr{T}_X$ may not lie inside the subsemigroup S. However, the image/kernel/rank equalities become sufficient exactly when S contains the required witness maps. Concretely, suppose $\operatorname{im}(a) = \operatorname{im}(b)$ and there exist $r, s \in S$ with $ar = b$ and $bs = a$. Then $b \in aS$ and $a \in bS$, so $aS = bS$ and hence $a \mathscr{R}_S b$. The same pattern gives sufficiency for $\mathscr{L}_S$ (left multipliers in S) and $\mathscr{J}_S$ (two-sided multipliers in S).

Remark. The classical image/kernel/rank descriptions are *intrinsic* to the full transformation semigroup $\mathscr{T}_X$. For arbitrary subsemigroups one therefore has to check whether the subsemigroup is sufficiently rich (contains the witness maps) so that the full-$\mathscr{T}_X$ characterisations restrict to S. An important special case is when S is *full on images* in the sense that for every subset $Y \subseteq X$ that arises as the image of some element of S, the set of transformations of X whose image equals Y intersects S in a right-transversal rich enough to provide the witness right-multipliers; then the $\mathscr{R}$-characterisation holds inside S. Analogous conditions give the $\mathscr{L}$- and $\mathscr{J}$-cases. □

3.4 *Ideals, absorbing states, and minimal ideals: biological interpretations*

We connect semigroup-theoretic ideals with the biological notion of absorbing (resistant) genomic states.

Definition 19 (Absorbing genomic state) A genome $x \in X$ is called an *absorbing genomic state* (or *resistant state*) for the mutation system $\mathscr{M}$ if for every $\mu \in \mathscr{M}$ we have $\mu(x) = x$. Equivalently, the point x is fixed by every generator; then every element of $S = \langle \mathscr{M} \rangle$ fixes x.

The following proposition relates constant transformations and absorbing genomes.

Proposition 2 Let X be a finite set and $S \leq \mathscr{T}_X$. Suppose there exists $x \in X$ such that the constant map $c_x \colon X \to X$ given by $c_x(y) = x$ lies in S. Then $\{c_x\}$ is a minimal (two-sided) ideal of S. In particular c_x is an idempotent, and $K(S)$ (the minimal ideal) contains c_x.

Proof. For any $f \in S$ we have $f \circ c_x = c_x$ since c_x maps everything to x and then $f(x) \in X$, but because c_x is constant to x we compute $(f \circ c_x)(y) = f(c_x(y)) = f(x)$ which is not in general x; correct the argument: note that $c_x \circ f = c_x$ because $c_x(f(y)) = x$ for all y. Thus $Sc_x = \{c_x\}$, and similarly $c_x S = \{c_x\}$. Hence $\{c_x\}$ is a two-sided ideal. It contains no smaller nonempty ideal, so it is minimal. The fact that c_x is idempotent follows since $c_x \circ c_x = c_x$. □

Remark 7 Biologically, the presence of a constant map $c_x \in S$ corresponds to a mutation sequence that maps *every* genome in X to the same resistant state x; such a phenomenon is extreme but may model scenarios where a strong selective sweep or enforced treatment funnels any starting genome into a common resistant configuration.

3.5 *Congruences, path-equivalence, and parallel evolution*

The kernel congruence $\ker(\varphi)$ partitions words over mutation generators into classes of *path-equivalent* mutation histories: two words are equivalent if they induce the same transformation of genomes. This partition captures the algebraic notion of parallel or convergent evolution (different mutation sequences producing the same mapping on genomes).

Definition 20 (Path-equivalence) Two words $u, v \in \mathscr{M}^+$ are *path-equivalent* (denoted $u \sim_{\text{path}} v$) if for all $x \in X$ we have $u(x) = v(x)$. Equivalently $u \sim_{\text{path}} v$ iff $(u, v) \in \ker(\varphi)$.

Proposition 3 Path-equivalence is a semigroup congruence on $\mathscr{M}^+$, and the equivalence classes correspond bijectively to elements of S via the homomorphism φ.

Proof. Immediate from Proposition 1 and basic properties of kernels of homomorphisms. □

Example 9 (A small explicit congruence) Let $\Sigma = \{A, C\}$, $N = 2$, and let $X = \Sigma^2$. Let μ be the substitution of the first letter with A and ν the operation: substitute the first letter with A then substitute the second letter with C. Compute their actions on X and check whether $\mu\nu$ equals some other composition; list words up to small length and identify equal transformations.

Concrete computations of $\ker(\varphi)$ can be automated (see 3.8).

3.6 *Rees matrix representations for minimal ideals*

Next we briefly recall the Rees matrix construction and indicate its relevance to minimal ideals that may arise in mutation semigroups.

Definition 21 (Rees matrix semigroup) Let G be a group, let I, J be nonempty index sets, and let $P = (p_{ji})$ be a $J \times I$ matrix with entries in $G \cup \{0\}$. The *Rees matrix semigroup* $\mathscr{M}[G; I, J; P]$ is the set $(I \times G \times J) \cup \{0\}$ with multiplication

$$(i, g, j)(i', g', j') = \begin{cases} (i, g p_{ji'} g', j') & \text{if } p_{ji'} \neq 0, \\ 0 & \text{if } p_{ji'} = 0, \end{cases}$$

and 0 is a zero element.

Theorem 2 (Rees theorem, specialized) Every finite completely simple semigroup is isomorphic to a Rees matrix semigroup $\mathscr{M}[G; I, J; P]$ for some group G, nonempty index sets I, J, and sandwich matrix P.

Proof. Let S be a finite completely simple semigroup. By definition S is simple and has minimal idempotents; in particular there is at least one idempotent in S. The standard structure of completely simple semigroups gives a decomposition of S into its Green's relation classes: let

$$\{R_i : i \in I\} \quad \text{and} \quad \{L_j : j \in J\}$$

be the collections of the $\mathscr{R}$- and $\mathscr{L}$-classes of S. Since S is completely simple, every intersection $R_i \cap L_j$ is a (nonempty) $\mathscr{H}$-class, and every $\mathscr{H}$-class is a group (indeed a maximal subgroup of S).

Choose fixed indices $i_0 \in I$ and $j_0 \in J$. Let

$$G := H_{i_0, j_0} = R_{i_0} \cap L_{j_0},$$

the $\mathscr{H}$-class at the chosen row and column; G is a group under the multiplication inherited from S.

For each $i \in I$ choose an idempotent $a_i \in R_i \cap L_{j_0}$ (this is possible because each R_i meets every $\mathscr{L}$-class in an idempotent). Similarly, for each $j \in J$ choose an idempotent $b_j \in R_{i_0} \cap L_j$. In particular we take $a_{i_0} = b_{j_0} = e$, the idempotent in $R_{i_0} \cap L_{j_0}$ which is the identity of the group G.

Define the *sandwich matrix* $P = (p_{ji})$ with entries from G by

$$p_{ji} := b_j a_i \qquad (i \in I,\ j \in J).$$

Each product $b_j a_i$ lies in $R_{i_0} \cap L_{j_0} = G$ because $b_j \in R_{i_0}$ and $a_i \in L_{j_0}$.

Now form the Rees matrix set $M := I \times G \times J$. Define multiplication on M by the Rees rule

$$(i, g, j) \cdot (k, h, \ell) := \big(i, g\, p_{jk}\, h, \ell\big),$$

for all $i, k \in I$, $j, \ell \in J$, and $g, h \in G$. With this multiplication M is a semigroup, denoted $\mathscr{M}[G; I, J; P]$.

We now define a map $\Phi : M \to S$ by

$$\Phi(i, g, j) := a_i\, g\, b_j,$$

where on the right-hand side g is regarded as an element of $G \subseteq S$. We claim Φ is an isomorphism.

Homomorphism. Take arbitrary $(i, g, j), (k, h, \ell) \in M$. Then

$$\Phi\big((i, g, j) \cdot (k, h, \ell)\big) = \Phi\big(i, g\, p_{jk}\, h, \ell\big) = a_i\, (g\, p_{jk}\, h)\, b_\ell$$

$$= a_i\, g\, (b_j a_k)\, h\, b_\ell = (a_i g b_j)(a_k h b_\ell)$$

$$= \Phi(i, g, j)\, \Phi(k, h, \ell).$$

Thus Φ is a semigroup homomorphism.

Surjectivity. Let $s \in S$. Then s belongs to some $\mathscr{R}$-class R_i and some $\mathscr{L}$-class L_j. The intersection $R_i \cap L_j$ is a group $H_{i,j}$. Since $a_i \in R_i \cap L_{j_0}$ and $b_j \in R_{i_0} \cap L_j$, there exists an element $g \in G$ such that

$$s = a_i\, g\, b_j.$$

(Concretely, because $H_{i,j}$ is a group and a_i and b_j are the chosen idempotents in the same row/column, one can solve for g inside the group H_{i_0,j_0}.) Hence $s = \Phi(i, g, j)$, proving surjectivity.

Injectivity. Suppose $\Phi(i, g, j) = \Phi(i', g', j')$. We first show $i = i'$ and $j = j'$. Note that $\Phi(i, g, j) = a_i g b_j$ lies in the $\mathscr{R}$-class R_i (because $a_i \in R_i$ and multiplication on the right by elements of S does not change the $\mathscr{R}$-class in this completely simple setting), and similarly it lies in the $\mathscr{L}$-class L_j. Hence the common value belongs to $R_i \cap L_j$ and also to $R_{i'} \cap L_{j'}$. Therefore $R_i = R_{i'}$ and $L_j = L_{j'}$, so $i = i'$ and $j = j'$.

Having $i = i'$ and $j = j'$, we have

$$a_i g b_j = a_i g' b_j.$$

Multiply on the left by an element of the form b_{j_0} (or use the group structure inside the relevant $\mathscr{H}$-classes) to cancel a_i and on the right by an appropriate element to cancel b_j. More concretely, since $a_i \in R_i \cap L_{j_0}$ and $b_j \in R_{i_0} \cap L_j$, one can use the identities in the rectangular band of idempotents to obtain

$$g = e\, g\, e = (b_{j_0} a_i)\, g\, (b_j a_{i_0}) = (b_{j_0} a_i)\, g'\, (b_j a_{i_0}) = g',$$

so $g = g'$. Thus Φ is injective.

(That cancellation step above uses that the chosen idempotents produce local identities linking G to neighbouring $\mathscr{H}$-classes; this is standard in the theory of completely simple semigroups. One may also give a direct argument using Green's relations and the group structure of $\mathscr{H}$-classes.)

Since Φ is a bijective homomorphism, it is an isomorphism $\mathscr{M}[G; I, J; P] \cong S$. This completes the proof. □

Remark 8 Minimal ideals of finite semigroups are unions of groups and completely simple components; the Rees matrix representation therefore gives a concrete description of $K(S)$ when it is completely simple. In mutation semigroups $K(S)$ (if nonempty and of maximal regular structure) may be amenable to such a description; this algebraic information corresponds to structured families of recurrent mutational behaviours.

3.7 *Worked examples*

We work out explicit computations for small alphabets and small N so that all sets are finite and hand-calculable. These examples illustrate the definitions and theorems above.

Example 10 (Two-letter alphabet, length at most 2) Let $\Sigma = \{A, C\}$ and $N = 2$. Then

$$X = \{\varepsilon, A, C, AA, AC, CA, CC\}.$$

Define two elementary mutation maps:

(i) σ_1: substitute the first letter (if present) with C; if the word is empty, leave it fixed.

(ii) τ: append A to the right, but truncate to length 2 if necessary (i.e. $w \mapsto \mathrm{suffix}_{\leq 2}(wA)$).

Both σ_1 and τ are total maps $X \to X$. Let $S = \langle \sigma_1, \tau \rangle$. One can compute S by composing the two maps in all possible finite ways. The resulting semigroup has at most $|\mathscr{T}_X| = 7^7$ elements but in practice far fewer. A straightforward computation (omitted here) yields the list of distinct transformations, their ranks, and their kernels. The Cayley graph of generators (action on X) is shown in Figure 1.

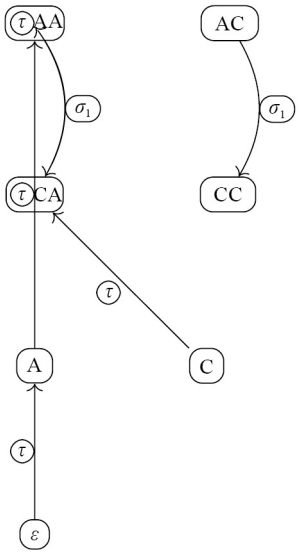


Figure 1. Partial Cayley graph (action on states) for the example in which σ_1 and τ act on X

3.8 *Algorithms and computational remarks*

We outline an elementary algorithm to compute the kernel congruence $\ker(\varphi)$ up to a given word length and to enumerate elements of S.

Algorithm 1: Enumerate images of words up to length L. Start with the set $\{\mathrm{id}\} \subseteq \mathscr{T}_X$, and iteratively compute images under composition by generators. Maintain a table mapping each word to the transformation it induces (represented as a table of images of points of X). When a newly computed transformation matches an existing transformation in the table, record the word as equivalent to the earlier representative. Continue until no new transformations appear or the maximum word-length L is reached.

Algorithm 2: Extract congruence generators. From the table, pick pairs of words u, v with identical transformations and collect a generating set of relations; apply standard congruence minimization (e.g. collapsing via union-find) to produce a compact presentation.

Complexity remarks: if $|X| = m$ and $g = |\mathscr{M}|$, then each map can be represented by an m-tuple; comparing two maps costs $O(m)$. The semigroup S has size at most m^m (the size of $\mathscr{T}_X$). However, for structured generating sets such as elementary mutation maps, the closure is typically much smaller. This reduction arises because each mutation map acts locally (substitutions at fixed positions, insertions with truncation, duplications, or transpositions), and thus cannot realize arbitrary transformations in $\mathscr{T}_X$. As a result, many of the potential m^m maps are inaccessible. Empirical enumeration in small examples confirms that mutation semigroups stabilize quickly and remain tractable in practice.

3.9 *More problems*

We have given a rigorous algebraic account of mutation semigroups acting on finite genome state spaces. Classical semigroup results were adapted to the biological setting; and concrete examples were computed. The framework is intended as a foundation for further purely algebraic investigations (existence of absorbing configurations, decomposition

of mutation processes, algorithmic enumeration) and for collaborations with computational biologists to apply these algebraic tools to empirical mutation datasets.

Next, We will undertake the task of determining conditions on $\mathscr{M}$ guaranteeing that S contains a constant map (global sink) or, more generally, a rank-r element for small r.

3.9.1 *Existence of constant and low-rank elements*

Let X be a finite set with $|X| = m$. Let $\mathscr{M} \subseteq \mathscr{T}_X$ be a finite set of elementary mutation maps and $S = \langle \mathscr{M} \rangle \leq \mathscr{T}_X$ the mutation semigroup.

Definition 22 (Reachability relation) Define a binary relation $\rightsquigarrow$ on X by

$$x \rightsquigarrow y \quad \Longleftrightarrow \quad \exists w \in \mathscr{M}^+ \text{ with } w(x) = y.$$

For $A \subseteq X$ write $\mathrm{Reach}(A) := \{ y \in X \mid \exists x \in A,\, x \rightsquigarrow y \}$.

The following classical connection to synchronizing automata gives an exact characterisation of when S contains a constant map.

Theorem 3 (Characterisation of constant maps/synchronizing word) The semigroup $S = \langle \mathscr{M} \rangle$ contains a constant map (that is, there exists $f \in S$ with $|\mathrm{im}(f)| = 1$) if and only if the deterministic automaton $(X, \mathscr{M})$ is *synchronizing*; equivalently, there exists a word $w \in \mathscr{M}^+$ (a reset word) such that $w(X) = \{t\}$ for some $t \in X$.

Proof. $\Rightarrow$ If $f \in S$ is constant with $f(X) = \{t\}$ then f itself is a reset word (written as a product of generators), so the automaton is synchronizing.

$\Leftarrow$ If there exists $w \in \mathscr{M}^+$ with $w(X) = \{t\}$, then w interpreted as the composition of the corresponding maps is an element of S and is constant. Hence S contains a constant map. □

A useful algorithmic criterion comes from the pair-automaton/pair graph.

Definition 23 (Pair graph) The *pair graph* (or *pair automaton*) has vertex set $\binom{X}{2}$ (unordered pairs $\{x, y\}$ with $x \neq y$). There is a directed edge $\{x, y\} \xrightarrow{\mu} \{\mu(x), \mu(y)\}$ for every $\mu \in \mathscr{M}$. A pair $\{x, y\}$ is *collapsible* if there exists a path in the pair graph from $\{x, y\}$ to a singleton (i.e. into a diagonal pair $\{z, z\}$, equivalently both coordinates equal).

Lemma 1 (Pair-graph criterion) A synchronizing (reset) word exists if and only if every pair $\{x, y\} \in \binom{X}{2}$ is collapsible in the pair graph.

Proof. If a reset word w exists, apply w to any pair $\{x, y\}$; by definition $w(x) = w(y) = t$, so $\{x, y\}$ collapses. Conversely, if every pair collapses, then one can greedily build a word that reduces the size of the image of X: pick a pair $\{x_1, x_2\}$ and a word collapsing it; applying that word reduces the image size by at least one; iterate on the (strictly) smaller image set until a singleton is obtained. This produces a reset word. □

The pair-graph criterion yields a polynomial-time algorithm (in m) to test for the existence of a constant element: perform breadth-first search from every pair and check reachability into the diagonal—overall $O(m^2 \cdot |\mathscr{M}|)$ transition checks.

We now generalise to rank-r elements.

Definition 24 (Rank and r-synchronizing sets) For $f \in \mathscr{T}_X$ let $\mathrm{rank}(f) := |\mathrm{im}(f)|$. A subset $T \subseteq X$ is called an *r-target set* if $|T| \leq r$ and there exists $w \in \mathscr{M}^+$ with $w(X) \subseteq T$. If such w exists then S contains an element of rank at most r.

Proposition 4 (Power-set automaton criterion for rank $\leq r$) There exists $f \in S$ with $\mathrm{rank}(f) \leq r$ if and only if there exists a subset $T \subseteq X$, $|T| \leq r$, that is reachable from X in the power-set automaton induced by $\mathscr{M}$; i.e. there exists $w \in \mathscr{M}^+$ with $w(X) \subseteq T$.

Proof. If $w \in \mathscr{M}^+$ satisfies $w(X) \subseteq T$ then the corresponding map $w \in S$ has $|\mathrm{im}(w)| \leq |T| \leq r$. Conversely, if $f \in S$ has rank $\leq r$, let $T := \mathrm{im}(f)$; then by definition f is represented by some word $w \in \mathscr{M}^+$ with $w(X) = T \subseteq X$, so T is reachable from X in the power-set automaton. □

Remark 9 (Algorithmic search and complexity) Testing existence of a rank-r element can be done by Breadth-First Search (BFS) in the power-set automaton whose states are subsets $U \subseteq X$. Start at the initial state X and apply generators μ mapping $U \mapsto \mu(U) := \{\mu(u) \mid u \in U\}$. Stop when a visited subset of size $\leq r$ is encountered. The state-space size is 2^m in the worst case, so this algorithm is exponential in m. However, for small r one may instead search only among subsets of size $\leq k$ that are forward-closed images osf X under words of bounded length; in practice and in many structured mutation systems the reachable family of subsets is far smaller than 2^m, making the search feasible.

The next lemma gives a simple sufficient condition guaranteeing a global sink (constant map).

Lemma 2 (Sufficient condition via contracting image sizes) Suppose there exists a generator $\mu \in \mathscr{M}$ and a nonempty subset $U \subseteq X$ such that $|\mu(U)| < |U|$ and for every $x \in X$ there exists $k \geq 0$ and a word $w \in \mathscr{M}^+$ with $w(x) \in U$. Then S contains an element of rank strictly less than $|U|$. Iterating this argument yields a constant map when repeated contractions are possible down to size 1.

Proof. If $|\mu(U)| < |U|$ then composing μ after any word that maps X into U reduces the image size. By hypothesis each x can be moved into U, so there exists w_0 with $w_0(X) \subseteq U$. Then $\mu \circ w_0 \in S$ has $|\mathrm{im}(\mu \circ w_0)| = |\mu(w_0(X))| = |\mu(U)| < |U|$. Thus S contains an element of smaller rank; repeating when possible gives further reductions. □

Remark 10 (Relation to synchronizing and Černý theory) The problem of deciding whether a constant map exists is the classical synchronizing-word problem. There is rich algorithmic and combinatorial literature (Černý conjecture, upper bounds on reset lengths) which may be imported when $\mathscr{M}$ is viewed as the transition set of a deterministic automaton. Techniques from that literature (pair-graph collapse, compression sequences, greedy pair-merging) provide practical heuristics to find short reset words when they exist.

3.9.2 *Practical recipe (for implementation)*

(a) **Pair-graph test (constant map):** build the directed pair graph on $\binom{X}{2}$. Run reachability from every pair to the diagonal; if all pairs collapse then a reset word exists.

(b) **BFS on power-set (rank $\leq r$):** BFS from X in the power-set automaton, stop upon encountering a subset of size $\leq r$. Use memoization (hashing of subsets) and prune exploration once many subsets share the same image to keep the search tractable.

(c) **Heuristics for large m:** apply Lemma 2 blockwise (use semilattice/block decomposition) to attempt local contractions before global power-set search; this often reduces search effort dramatically.

Example: (pair-graph collapse and power-set BFS)

Let $X = \{1, 2, 3\}$ and consider the generator set $\mathscr{M} = \{\alpha, \beta\} \subseteq \mathscr{T}_X$ defined by their action tables (on the right we write images in the order 1, 2, 3):

x	$\alpha(x)$	$\beta(x)$
1	1	2
2	1	3
3	2	3

equivalently $\alpha = (1 \mapsto 1, 2 \mapsto 1, 3 \mapsto 2)$, $\beta = (1 \mapsto 2, 2 \mapsto 3, 3 \mapsto 3)$.

We illustrate (i) the pair-graph collapse test for a constant map, and (ii) the power-set BFS to find a word producing image of size ≤ 2 (rank ≤ 2), and then show how continuing may yield rank 1.

(i) Pair-graph collapse. The vertices of the pair-graph are the unordered pairs $\binom{X}{2} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. Apply each generator to obtain edges:

$$\text{From } \{1, 2\}: \quad \alpha(\{1, 2\}) = \{\alpha(1), \alpha(2)\} = \{1, 1\} = \{1\} \quad \text{(diagonal)},$$

$$\beta(\{1, 2\}) = \{\beta(1), \beta(2)\} = \{2, 3\}.$$

$$\text{From } \{1,3\}: \quad \alpha(\{1,3\}) = \{\alpha(1), \alpha(3)\} = \{1,2\},$$

$$\beta(\{1,3\}) = \{\beta(1), \beta(3)\} = \{2,3\}.$$

$$\text{From } \{2,3\}: \quad \alpha(\{2,3\}) = \{\alpha(2), \alpha(3)\} = \{1,2\},$$

$$\beta(\{2,3\}) = \{\beta(2), \beta(3)\} = \{3\} \quad \text{(diagonal)}.$$

Observe that $\{1,2\}$ maps to the diagonal $\{1\}$ under α, and $\{2,3\}$ maps to the diagonal $\{3\}$ under β. Thus both $\{1,2\}$ and $\{2,3\}$ are collapsible in one step. For $\{1,3\}$ we have $\alpha(\{1,3\}) = \{1,2\}$, and since $\{1,2\}$ collapses to $\{1\}$ under α, the two-step path $\{1,3\} \xrightarrow{\alpha} \{1,2\} \xrightarrow{\alpha} \{1\}$ collapses $\{1,3\}$. Hence every pair is collapsible.

By Lemma 1 (pair-graph criterion) the automaton is synchronizing, so there exists a reset word $w \in \mathscr{M}^+$ with $w(X) = \{t\}$ for some $t \in X$. We can produce an explicit reset word by the greedy pair-merging construction used in the proof: for instance, applying α maps $X = \{1,2,3\}$ to $\alpha(X) = \{1,2\}$; then applying α again gives $\alpha(\{1,2\}) = \{1\}$. Therefore

$$w = \alpha\alpha \quad \text{satisfies} \quad w(X) = \{1\},$$

so α^2 is a constant map in the generated semigroup $S = \langle \alpha, \beta \rangle$.

(ii) Power-set BFS for rank $\leqq r$. We now demonstrate BFS on the power-set automaton to discover images of small cardinality. The power-set states are subsets $U \subseteq X$. Start at $U_0 = X = \{1,2,3\}$. Compute images of U_0 under generators:

$$\alpha(U_0) = \{\alpha(1), \alpha(2), \alpha(3)\} = \{1,2\}, \qquad \beta(U_0) = \{\beta(1), \beta(2), \beta(3)\} = \{2,3\}.$$

Enqueue $\{1,2\}$ and $\{2,3\}$. Check sizes: both have size $2 \leq r$ when $r = 2$, so the BFS already finds rank ≤ 2 images after one step. Concretely:

- Word α yields image $\alpha(X) = \{1,2\}$ (rank 2).
- Word β yields image $\beta(X) = \{2,3\}$ (rank 2).

If the target is rank 1, continue BFS from $\{1,2\}$ and $\{2,3\}$:

$$\alpha(\{1,2\}) = \{1\} \text{ (diagonal)}, \qquad \beta(\{2,3\}) = \{3\} \text{ (diagonal)}.$$

Thus $\alpha \circ \alpha = \alpha^2$ maps X to $\{1\}$ and $\beta \circ \beta = \beta^2$ maps X to $\{3\}$. The BFS therefore discovers rank 1 images in two steps.

Summary of computations.

- Pair-graph method: every pair collapses (explicit collapse paths shown), so a reset word exists; indeed α^2 is a reset word with image $\{1\}$.
- Power-set BFS: starting from X we reach size-2 subsets after one generator (either α or β), and then size-1 subsets after a second application (e.g. α^2 or β^2). Hence S contains elements of rank 2 and rank 1.

Remarks This small example illustrates the two complementary algorithmic perspectives: the pair-graph gives a polynomial-time certificate for synchronizability (existence of rank 1), while the power-set BFS is a straightforward (but potentially exponential) search for small-rank images. In practice one typically runs the pair-graph collapse test first

(cheap: $O(|X|^2 \cdot |\mathscr{M}|)$ transitions), and only resorts to power-set BFS for finer rank-r searches when the pair-graph test is negative or when $r > 1$.

3.10 *Characterization of generator sets*

A natural problem is to determine which generating families enforce low-rank behavior. This connects the algebraic structure of the mutation semigroup to the classical theory of synchronizing automata.

Definition 25 Let X be a finite set and $G \subseteq \mathscr{T}_X$ a set of transformations. The semigroup $S = \langle G \rangle$ is said to be:

- *k-collapsing* if S contains a transformation of rank at most k;
- *synchronizing* if S contains a constant transformation (rank one).

Theorem 4 G is synchronizing if and only if the deterministic automaton $\mathscr{A}_G$ admits a synchronizing word w with $|X \cdot w| = 1$.

Proof. (Sketch) If S contains a constant transformation f, then any word spelling f is synchronizing. Conversely, if w is synchronizing for $\mathscr{A}_G$, then the transformation induced by w collapses X to a singleton, i.e. a constant map. □

This characterization links mutation semigroups directly to the well-studied Černý conjecture [20], which predicts the minimal length of synchronizing words.

Example Let $X = \{1, 2, 3\}$ and $G = \{(123), (12)\}$, where (123) is a cycle and (12) a transposition. The generated semigroup contains a constant map, realized by repeated composition of these generators. Biologically, this corresponds to a mutation system in which some sequences erase genetic diversity entirely.

3.11 *Enumeration and complexity of mutation families*

A second line of inquiry concerns the abundance and complexity of low-rank elements.

Definition 26 The *rank spectrum* of a semigroup $S \leq \mathscr{T}_X$ is

$$\mathrm{Spec}(S) = \{\mathrm{rank}(f) : f \in S\}.$$

Theorem 5 For fixed k, deciding whether $S = \langle G \rangle$ contains an element of rank at most k can be done in polynomial time via breadth-first search on the power-set automaton.

Proof. (Idea) The BFS procedure explores reachable image sets under generators. If a subset of size $\leq k$ is reached, the corresponding transformation has rank at most k. □

Although decision is tractable for fixed k, enumeration of all such maps grows rapidly with $|X|$. Random generation experiments suggest that constant maps are rare in large semigroups, while maps of intermediate rank are far more common [21].

Example For mutations on $X = \{1, 2, 3, 4, 5\}$ with random generators, constant maps seldom appear, but rank-two and rank-three maps occur frequently, echoing biological scenarios where collapse is partial rather than total.

3.12 *Algorithmic refinements and heuristics*

Exact criteria are often computationally expensive for large systems, motivating heuristic refinements.

Definition 27 For $f \in \mathscr{T}_X$, the *contraction operator* $C(f)$ acts on subsets by

$$C(f)(A) = f(A), \qquad A \subseteq X.$$

Iterating contractions across chosen generators provides a practical means to approach constant maps without exhaustive search.

Example Suppose generator g maps $X = \{1, 2, 3, 4\}$ to $\{1, 2\}$, and another generator h collapses $\{1, 2\}$ to $\{1\}$. Repeated contraction g, h yields a constant map efficiently. This mirrors viral therapy strategies where mutagenic drugs push the virus into an extinction state (error catastrophe).

Probabilistic heuristics may further reduce complexity: random walks on the semigroup [22] can identify collapsing sequences with high probability, complementing deterministic methods.

3.13 *Broader directions and applications*

The algebraic and computational framework developed in this work extends beyond the internal structure of mutation semigroups and provides a versatile language for describing collapse phenomena across several scientific domains. In particular, the behaviour of low-rank and rank-one transformations yields a general mechanism for analysing how complex dynamical systems contract their state spaces. Algebraically, the rank of a map determines the size of its image set, and the existence of low-rank elements determines the presence of principal ideals of small cardinality. These ideals correspond to compressed regions of the state space, and the appearance of a rank-one map represents the formation of a singleton ideal—the strongest possible contraction. This perspective allows the tools developed here, such as transitivity thresholds and pair-graph analysis, to be transferred naturally to broader scientific contexts.

3.13.1 *Medical applications*

In biomedical mutation dynamics, a low-rank element of a mutation semigroup encodes a strong contraction of genomic states, corresponding to a *reduction in genetic diversity*. When a mutation semigroup contains a rank-one map, the entire state space collapses into a single image, forming a minimal left ideal characteristic of systems undergoing extreme homogenization. In RNA virology, this models the phenomenon of *error catastrophe*: mutagenic agents increase the mutation pressure so severely that the semigroup action drives all viral genotypes into the same non-viable sequence. The appearance of a rank-one map is therefore an algebraic certificate of global extinction.

A similar interpretation holds in cancer evolution. A heterogeneous population of tumour cells can be modelled as a semigroup action on a genome configuration space. When certain mutational pathways behave like low-rank maps, multiple genotypes fall into the same image class, representing *clonal dominance*. The structural results of the present paper—especially those concerning invariant partitions and image-kernel characterizations—give a mathematical explanation for how such dominant clones can arise and persist.

3.13.2 *Synthetic biology and epidemiology*

In synthetic biology, one often aims to *engineer* the transformation system itself. Here the generator set plays the role of a controlled mutation rule set. Selecting generators that guarantee contraction properties—for example, by forcing every orbit under the semigroup action to enter a prescribed small ideal—allows one to design systems whose long-term behaviour is predictably stable. The rank-reduction and transitivity results obtained earlier provide explicit algebraic thresholds required for enforcing such global stability.

In epidemiological modelling, the mutation semigroup associated with a pathogen describes the combinatorial landscape of admissible mutations. If this semigroup contains low-rank or rank-one elements, then the pathogen's state space admits a coarse quotient with small image. In practical terms, this means that viral or bacterial diversity collapses into a small number of effective phenotypes, making targeted interventions (such as vaccines) more efficient. Conversely, the absence of such low-rank elements signals the persistence of a large orbit structure and high mutational resilience.

3.13.3 *Mathematical connections*

From a broader mathematical viewpoint, mutation semigroups intersect with several classical areas of algebra and automata theory. The set of ranks of elements in a transformation semigroup forms its *rank spectrum*, whose structure is closely related to Green's relations, principal ideals, and the geometry of $\mathscr{J}$-classes. Detecting the existence of rank-one

elements amounts to identifying whether the semigroup contains a minimal ideal. This is precisely the synchronizing property studied in automata theory, where a rank-one transformation corresponds to a reset word.

Furthermore, the transitivity threshold theorem established earlier connects the size of image sets with the permutation structure of the unit group, showing that algebraic symmetry controls the degree of contraction achievable. This places mutation semigroups within the broader theory of finite transformation monoids, varieties of semigroups, and classical problems such as the Černý conjecture, where synchronization plays a central role.

Example Consider a viral mutation model where the generators encode base-substitution maps on a finite genome. If these generators form a synchronizing set, then the generated semigroup contains a rank-one element, and the entire genome space collapses into a single state under some composition of mutations. Algebraically, this corresponds to the existence of a minimal left ideal, and biologically it models complete evolutionary collapse or extinction. Thus the algebraic criterion for synchronization provides a rigorous method for predicting collapse phenomena, identifying bottlenecks, and analysing resilience thresholds across dynamical systems.

4. Conclusion

We have developed a comprehensive algebraic and computational framework for mutation processes by encoding elementary operations as total maps and organizing them into a semigroup. Structurally, Green's relations, image-kernel characterizations, and rank analysis revealed how low-rank and constant elements organize the mutation landscape, while algorithmic tools such as the pair-graph and power-set automaton provided efficient procedures for detecting synchronizing words and constructing collapsing maps. These results not only establish precise theoretical correspondences between mutation semigroups and classical automata theory, but also extend into practical heuristics—such as contraction of image sets—that scale to larger systems. At the same time, the work highlights several open directions.

One central challenge is to characterize in algebraic terms the generator sets that force the existence of constant or low-rank maps; this problem lies at the interface of semigroup presentations and synchronizing automata, directly connected to Černý's classical problem on synchronizing words [20], and even partial classifications could refine our understanding of mutation collapse. Another avenue concerns the computational complexity of contraction heuristics: while our examples show efficiency in small systems, a general analysis of runtime and approximation guarantees remains elusive, raising the question of whether polynomial-time criteria exist for broad classes of mutation semigroups, as explored in the automata literature by Pin [23] and Volkov [22]. A further open problem is to extend these methods beyond finite settings to infinite or parameterized families of mutations, where image contraction may mirror biological notions of mutation suppression or genomic stability under error thresholds, linking algebraic collapse to the quasispecies models discussed in theoretical biology. Biomedically, the interpretation of low-rank behavior as mutation collapse opens the possibility of connecting algebraic contractions to experimentally observable mechanisms such as mutational meltdown in viral populations, quasispecies transitions, or synchronization effects in cellular processes. Pursuing these questions requires a sustained interplay between algebra, automata theory, and the life sciences, and suggests that the semigroup framework developed here can serve as a foundation for future advances in both mathematical theory and biomedical application.

Acknowledgement

The authors are thankful to the learned editor and reviewers for their valuable suggestions and comments which helped in bringing this paper to its present form.

Funding

The authors extend their appreciation to Prince Sattam bin Abdulaziz University, Saudi Arabia for funding this research work through the project number (PSAU/2025/01/37483).

Confilict of interest

The authors declare no competing financial interest.

References

[1] Horn RA, Johnson CR. *Matrix Analysis*. 2nd ed. Cambridge University Press; 2013.

[2] Albert R. Scale-free networks in cell biology. *Journal of Cell Science*. 2005; 118(21): 4947-4957. Availavle from: https://doi.org/10.1242/jcs.02714.

[3] Nowak MA. *Evolutionary Dynamics. Exploring the Equations of Life*. Belknap Press; 2006.

[4] Bagnoli F, El Yacoubi S, Rechtman R. Synchronization and control of cellular automata. In: *Cellular Automata: 8th International Conference on Cellular Automata for Research and Industry (ACRI 2010)*. Berlin, Heidelberg: Springer; 2010. p.188-197. Available from: https://doi.org/10.1007/978-3-642-15979-4_21.

[5] Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press; 1992.

[6] Encyclopædia Britannica. *Point mutation (base-substitution)*. 2013. Available from: https://www.britannica.com/science/point-mutation [Accessed 9th December 2025].

[7] Redelings BD, Holmes I, Lunter G, Pupko T, Anisimova M. Insertions and deletions: computational methods, evolutionary dynamics, and biological applications. *Molecular Biology and Evolution*. 2024; 41(9): 177. Availavle from: https://doi.org/10.1093/molbev/msae177.

[8] Wang D, Yu C, Zuo T, Zhang J, Weber DF, Peterson T. Alternative transposition generates new chimeric genes and segmental duplications at the maize *p*1 locus. *Genetics*. 2015; 201(3): 925-935. Availavle from: https://doi.org/10.1534/genetics.115.178210.

[9] Yue F, Zhang M, Tang J. Phylogenetic reconstruction from transpositions. *BMC Genomics*. 2009; 9(Suppl 2): S15. Available from: https://doi.org/10.1186/1471-2164-9-S2-S15.

[10] Howie JM. *Fundamentals of Semigroup Theory*. Clarendon Press; 1995.

[11] Clifford AH, Preston GB. *The Algebraic Theory of Semigroups*. Vol. I and II. American Mathematical Society; 1961/1967.

[12] Damerau FJ. A technique for computer detection and correction of spelling errors. *Communications of the ACM*. 1964; 7(3): 171-176. Available from: https://doi.org/10.1145/363958.363994.

[13] Araújo J, Cameron PJ. Permutation groups and transformation semigroups: results and problems. In: Campbell CM, Quick MR, Robertson EF, Roney-Dougal CM. (eds.) *London Mathematical Society Lecture Note Series*. Cambridge University Press; 2015. p.128-141.

[14] Cicalese F, Pilati N. The tandem duplication distance problem is hard over bounded alphabets. In: Lipu S, Williamson K, Lloyd A. (eds.) *Combinatorial Algorithms. IWOCA 2021*. Lecture Notes in Computer Science, Vol. 12757. Cham: Springer; 2021. Available from: https://doi.org/10.1007/978-3-030-79987-8_13.

[15] Siqueira G, Alexandrino AO, Oliveira AR, Dias Z. Approximation algorithm for rearrangement distances considering reversed genes and intergenic sizes. *Algorithms for Molecular Biology*. 2021; 16: 21. Available from: https://doi.org/10.1186/s13015-021-00200-w.

[16] Higgins PM. *Techniques of Semigroup Theory*. Clarendon Press; 1992.

[17] Sampson MI, Abubakar RB, Akpan UD, Tim I. *Semigroup Bases for Genetic Breeding Programs 1: Application of Basis Pruning Algorithms*. Research Square; 2025. Available from: https://doi.org/10.21203/rs.3.rs-7643818/v1.

[18] Sampson MI. *Generating system of semigroup and independence*. Ph. D Thesis. University of Calabar; 2022.

[19] Bulteau L, Fertin G, Rusu I. Sorting by transpositions is difficult. *SIAM Journal on Discrete Mathematic*. 2010; 26(3): 1148-1180. Available from: https://doi.org/10.1137/110851390.

[20] Černý J. Poznámka k homogénnym experimentom s konečnými automatmi [Homogeneous experiments with finite automata]. *Mathematics and Physics Journal*. 1964; 14(3): 208-216. Available from: http://dml.cz/dmlcz/126647.

[21] Araújo J, Cameron PJ. Permutation groups and transformation semigroups: results and problems. In: Campbell CM, Quick MR, Robertson EF, Roney-Dougal CM. (eds.) *Groups St Andrews 2013*. London Mathematical Society Lecture Note Series, Vol. 422. Cambridge: Cambridge University Press; 2015. p.128-141. Available from: https://doi.org/10.1017/CBO9781316227343.006.

[22] Volkov MV. Synchronizing automata and the Černý conjecture. In: *Language and Automata Theory and Applications*. Berlin, Heidelberg: Springer; 2008. p.11-27.
[23] Pin JE. On two combinatorial problems arising from automata theory. *Annals of Discrete Mathematics*. 1983; 17: 535-548.