# A New Model of Genetic Algorithm Using a Bipartite Graph and the Action of Largest Subgroup of Dihedral Group $D_n$ on Invariance Markov Basis, *n* is a Multiple of 6

# Husein Hadi Abbass<sup>1</sup>, Hussein Salman Mohammed Hussein<sup>2</sup>

Department of Mathematics, Faculty of Education for Girls, University of Kufa, Iraq

Abstract: In this paper, we introduce a new model of genetic algorithm that permutes the pieces of nucleotides in aligned DNA sequences using a bipartite graph and the action of largest subgroup **H** of dihedral Group  $D_n$ , *n* is multiple of 6 on  $\frac{n^2-3n}{3} \times 3 \times \frac{n}{3}$  - contingency tables with fixed two dimensional marginals and their Markov basis **B** such that **B** is **H**-invariant,. Where **B** is the Markov basis found by H. H. Abbass and H. S. Mohammed Hussein in [7].

Keywords: Computational algebraic statistics, sufficient statistics, bipartite graph, dihedral group, Markov basis

#### 1. Introduction

Since 1998 the publication of P.Diaconis and B.Sturmfels, the new field of computational algebraic statistics has been developing rapidly, and in the same year P.Diaconis and B.Sturmfels defined the notion of Markov basis for constructing a connected Markov chain for sampling from a conditional distribution over a discrete sample space and proved the fundamental fact that a Markov basis corresponds to a set of binomial generators of a toric ideal[11]. In 2000, M. Dyer, and C. Greenhill, found a Polynomial-time counting and sampling of two-rowed contingency tables[10]. In 2001, A.Dobra showed that the only moves that have to be included in a Markov basis that links all contingency tables having a set of fixed marginals when this set of marginals induces a decomposable graphical models[1]. In 2002, A. Dobra, and S. Sullivant, described a divide-and-conquer algorithm for generating Markov basis of multi-way tables that connects all tables of counts having a fixed set of marginal totals[2]. In 2003, S. Aoki and A.Takemura proved that there exists a unique minimal basis for 3×3×K contingency tables consisting of four types of indispensable moves [14], and in the same year S. Aoki, and A.Takemura presented a list of indispensable moves of unique minimal Markov basis of 3×4×K and 4×4×K contingency tables with fixed two- dimensional marginals[13], also A.Takemura, and S. Aoki. gave some characterizations of minimal Markov basis for connected Markov chain and given a necessary and sufficient condition for uniqueness of minimal Markov basis[3]. In 2005, A. Takemura, and S. Aoki. Studied the Markov basis for sampling from discrete sample space, which is equipped with some convent metric and they started from two state in the sample space, and they asked whether they can always move closer by an element of a Markov basis and they called a Markov basis distance reducing[4].

In [7] H. H. Abbass and H. S. Mohammed Hussein found a Markov basis **B** and toric ideals for  $\frac{n^2-3n}{3} \times 3 \times \frac{n}{3}$ -contingency tables with fixed two dimensional marginals, **n** is a multiple of 3 greater than or equal 6, also they [8] found the largest subgroup *H* of dihedral Group  $D_n$ , such that **B** is *H*-invariant, **n** is a multiple of 3.

In this paper, we use the Markov basis **B** and action of the subgroup H of dihedral Group  $D_n$  on these contingency tables to give a new model of permutation the pieces of nucleotides in DNA sequences.

#### 2. Preliminaries

In this section, we review some basic definitions and notations of contingency table, dihedral group, connected graph, bipartite graph, moves, Markov basis, and toric ideals that we need in our work.

**Definition 1(see[17]).**Let **1** be a positive integer greater than or equal<sup>3</sup>. The group of all symmetries of the regular polygon with n sides, including both rotations and reflections, is called **dihedral group** and denoted by  $D_{n}$ . If we center the regular polygon at origin then the elements of the dihedral group acts as linear transformation of the plane. Lets us represent the elements of  $D_n$  as matrix, with composition multiplication. Dihedral groups are among simplest examples of finite groups and they play an important role in group theory, geometry, and chemistry. The set of rotations is generated by r- counterclockwise rotation with angle  $2\pi / n$  of order **m**, and the set of reflections is of order 2 and every element splgenerates  $\{e, e\}$  $sr^{j}$ , where e is the identity element in Dn .  $D_{n}$  can be written as: $\{e, r, r^2, ..., r^{n-1}, s, sr, sr^2, ...,$  $Sr^{n-1}$ . In general, we can write  $D_n$ as:

 $\begin{array}{lll} D_n=\{s^jr^k: 0\leq k\leq n-1 \ , \ 0 \ \leq j \ \leq 1\} & \text{which has the} \\ \text{following properties:} & r^n=1, sr^ks=r^{-k}, \\ (sr^k)^2=1, \ for \ all \ 0 \ \leq k \ \leq n-1 \ . \text{The} \\ \text{composition of two elements of the } D_n & \text{is given by} \\ r^ir^j=r^{i+j}, \quad r^i \ s \ r^j=sr^{j-i}, \ sr^ir^j=sr^{i+j}, \\ sr^is \ r^j=r^{j-i}. \end{array}$ 

**Remark 1(see [17]).** If we label the vertices (of the regular n-gon) **1** to n in a counterclockwise direction around n-gon then the elements of  $D_n$  can be written as permutations of vertices, let r be a counterclockwise rotation, and let S be the reflection of the n-gon about an axis through the center and vertex **1**, as indicated in below. The element r generates the cyclic group of order n Cn which is a normal cyclic subgroup of  $D_n$ . In all cases, addition and subtraction should be performed using modular arithmetic with modulus n.



#### Elements of Cn Elements of Dn Any symmetry will fix the origin and is determined by the image of two adjacent vertices, say 1 and 2. The vertex 1 can be taken to any of n vertices and then the vertex 2 must be taken to one of the two vertices adjacent to the image of 1. Hence, $D_n$ is a non abelian group of order 2n generated by r and s.

Now, we give some concepts about the action of a group on a set that we use later.

**Definition 2**(see[12]). Let I be a finite set n = |I| elements, we call an element of I a cell and denoted by  $i \in I$ . i is often multi-index  $i = i_1 \dots i_m$ . A non-negative integer  $x_i \in \mathbb{N}$ denotes the frequency of a cell i. The set of frequencies is called *a contingency table* and denoted as  $x = \{x_i\}_{i \in I}$ , with an appropriate ordering of the cell, we treat a contingency table  $x = \{x_i\}_{i \in I} \in \mathbb{N}^n$  as a *n*-dimensional column vector of non-negative integers. Not that a contingency table can also be considered as a function from I to  $\mathbb{N}$  defined as  $i \mapsto x_i$ .

**Definition 3**(see [12]). The  $L_1$ -norm of  $\mathbf{x} \in \mathbb{N}^n$  is called the samplesize and denoted as

 $|\mathbf{x}| = \sum_{i \in I} x_i$  We will denote  $\mathbb{Z}$  be the set of integer numbers, also we denote to the  $a_j \in \mathbb{Z}^n, j = 1, ..., v$ , as fixed column vectors consisting of integers. A v-dimensional column vector  $\mathbf{t} = (t_1, ..., t_v)' \in \mathbb{Z}^v$  as  $t_j = a_j' \mathbf{x}, j = 1, ..., v$ . Here ' denotes the transpose of a vector or matrix. We also define a  $v \times p$  matrix A, with its *j*-row being  $a_j'$  given by  $A = \begin{bmatrix} a_1' \\ \vdots \\ a_v' \end{bmatrix}$ , and if  $\mathbf{t} = A\mathbf{x}$  is a v-

dimensional column vector, define the we set  $T = \{t: t = Ax, x \in \mathbb{N}^n\} = A\mathbb{N}^n \subset \mathbb{Z}^v$ . In typical situations of a statistical theory, t is *sufficient statistic* for the nuisance parameter. The set of **x**'s for а given t,  $A^{-1}[t] = \{x \in \mathbb{N}^n : Ax = t\} (t-\text{fibers}), \text{ is }$ considered for performing similar tests, for the case of the independence model of two-way contingency tables, for example, t is the row sums and column sums of  $\mathbf{x}$ , and  $A^{-1}[t]$  is the set of  $\mathbf{x}$ 's with the same row sums and column sums to t. The set of tfibers gives a decomposition of M<sup>n</sup>. An important observation is that *t*-fiber depends on given only through its kernel, ker (A). For different A's with the same kernel, the set of *t*-fibersare the same. In fact, if we define  $x_1 \sim x_2 \Leftrightarrow x_1 - x_2 \in \ker(A)$  this relation is an equivalence relation and M<sup>n</sup> is partitioned into disjoint equivalence classes. The set of *t*-fibers is simply the set of these equivalence classes. Furthermore, t may be considered as labels of these equivalence classes.

**Definition 4**(see [12]).A *n*-dimensional column vector of integers  $\mathbf{z} = \{z_i\}_{i \in I} \in \mathbb{Z}^n$  is called a move if it is in the kernel of A, i.e.  $A\mathbf{z} = \mathbf{0}$ .

Remark 2 (see[3]).For a movez, the positive part  $z^+ = \{z^+_i\}_{i \in I}$  and the negative part  $z^- = \{z^-_i\}_{i \in I}$  are  $z_{i}^{+} = \max(z_{i}, 0), z_{i}^{-} = \max(-z_{i}, 0),$ by defined  $z = z^{+} - z^{-}$ respectively, Then and  $z^+, z^- \in \mathbb{N}^n z^+, z^- \in \mathbb{N}^n$ . Moreover,  $z^+$  and  $z^-$  are in the same t-fiber, i.e.,  $\mathbf{z}^+, \mathbf{z}^- \in A^{-1}[t]$  for  $t = A\mathbf{z}^+ = A\mathbf{z}^-$ . We define the *degree* of  $\mathbf{z}$  as the sample size of  $\mathbf{z}^+$  or  $(\mathbf{z}^-)$  and denote it by deg(z) =  $|z^+| = |z^-|$ . In the following we denote the set of moves (for given A) а by  $M = M_A = \mathbb{Z}^n \cap \ker (A)$ .

**Definition 5** (see[12]).Let  $A:\mathbb{Z}^n \to \mathbb{Z}^v$  be a linear transformation,  $t \in \mathbb{Z}^v$ , and  $A^{-1}[t]$  be the set of t-fibers, and let  $B \subset \ker_{\mathbb{Z}}(A)$ , then we define  $A^{-1}[t]_B$  be the graph with vertex set  $A^{-1}[t]$  and u - -v an edge if and only if  $u - v \in \pm B$ .

**Definition 6** (see[3]).Let  $A^{-1}[t] = \{x \in \mathbb{N}^n : Ax = t\}$ . A set of finite moves **B** is called **Markov basis** if for all  $t, A^{-1}[t]$  constitutes one **B** equivalence class.

**Definition 7** (see[6]). A graph G is connected if for every pair of distinct vertices  $u, v \in V(G)$ , where V(G) be the set of vertices of the graph G, the graph G has a u, v-path. Otherwise, we say the graph is disconnected.

**Definition 8**(see[6]). A graph G is a bipartite graph if there are  $X, Y \subseteq V(G)$  meeting the following conditions:

 $1.V(G) = X \cup Y,$ 

 $2. X \cap Y = \Phi,$ 

3. G[X] and G[Y] are both null graphs, where G[X] and G[Y] are subgraphs of the graph G induced by the set of vertices  $X, Y \subseteq V(G)$  respectively.

# Volume 3 Issue 12, December 2014

Paper ID: OCT14464

www.ijsr.net Licensed Under Creative Commons Attribution CC BY Theorem 1(see[6]).

For a graph **G** the following statements are equivalent: 1. **G** is bipartite.

2. Every cycle in *G* has an even length.

**Definition 9 (see [6]).**Let  $B \subset M_A$  be the set of moves and let  $x_1, x_2 \in A^{-1}[t]$ . We say that  $x_2$  accessible from  $x_1$  by B if there exists a sequance of moves  $z_1, \ldots, z_k \in B$  and  $z_k \in \{-1, 1\}, k = 1, \ldots, K$ , such that

$$x_{2} = x_{1} + \sum_{k=1}^{K} \varepsilon_{k} z_{k},$$
  
$$x_{1} + \sum_{k=1}^{K} \varepsilon_{k} z_{k} \in A^{-1}[t] \text{ for } 1 \le k \le K.$$

**Definition 10**(see[3]). If  $B \subseteq ker_{\mathbb{Z}}(A)$  is a set such that  $A^{-1}[t]_{\mathbb{B}}$  is connected for all t, then B is a Markovbasis for A

**Remark 3:** Throughout this paper, the symbol  $\square$  denotes a field of complex numbers, the set  $\square^p$  is the vector space of p-tuples of elements in  $\square$ .

Henceforth  $P_1, P_2, \dots, P_n$  denote indeterminate, that is, polynomial variables. A monomial m in the indeterminates  $P_1, P_2, \dots, P_p$  is a expression of the form

 $m = \prod_{i=1}^{n} P^{\alpha_{i}}_{i}$ , where  $\alpha_{1}, \alpha_{2}, \dots, \alpha_{n}$ are nonnegative integers. We will often use the shorth  $m = \prod_{i=1}^{p} P^{\alpha_{i}} = P^{\alpha}$  to denote this monomial. shorthand A polynomial is a linear combination of finitely many CP monomials  $f(p) = f(P_1, P_2, ..., P_n) = \sum_{\alpha} c_{\alpha} P^{\alpha}$  where the  $c_{\alpha} \in \mathbb{C}$  and at most finitely many of them are nonzero. Note any polynomial f(P) is also a function from  $\mathbb{C}^{P}$  to  $\mathbb{C}$ , simply by evaluating the polynomial at a point of  $\mathbb{Q}^{\mathbb{P}}$ . The set of all polynomials in the **p** indeterminates  $P_1, P_2, \dots, P_p$  is denoted by either  $(P_1, P_2, \dots, P_p)$  or (P), for short. Note that (p) has the structure of a ring because we can add and multiply two polynomials to produce new polynomials, and these addition and multiplication operations are well-behaved with respect to one another.

**Definition 11** (see[16]). Let  $A:\mathbb{Z}^n \to \mathbb{Z}^d$  be a linear transformation, the toric ideal  $I_A$  is the ideal  $< P^u - P^v: u, v \in \mathbb{N}^n, A(u) = A(v) > \subseteq \mathbb{C}[P_1, \dots, P_p]$  where  $P^u = P_1^{u_1} P_2^{u_2} \dots P_p^{u_p}$ .

Remark 4 (see [7]). Let *n* be a multiple of 3 such that  $n \ge 6$ , and let  $x_j \in A^{-1}[t], j = 1, ..., k$  be the representative elements of the set of  $3 \times \frac{n}{3}$  -contingency tables and  $\mathbf{B} = \{\mathbf{z}_1, \mathbf{z}_2, ..., \mathbf{z}_k\}$  such that each  $\mathbf{z}_j$ , j = 1, 2, ..., k, is a matrix of dimension  $3 \times \frac{n}{3}$  either has two columns (1, -1, 0)', (-1, 1, 0)' ((1, 0, -1)', (-1, 0, 1)' or either (0, 1, -1)', (0, -1, 1)') and the other columns are zero denoted by  $+\mathbf{z}_j$ , or it two columns (1, -1, 0)', (-1, 0, 1)' or either (0, 1, -1)', (0, -1, 1)') and the other columns are zero denoted by  $+\mathbf{z}_j$ , or it two columns (1, -1, 0)', (-1, 1, 0)' ((1, 0, -1)', (0, -1, 1)') and the other columns are zero denoted by  $+\mathbf{z}_j$ , or it has two

columns  $(-1,0,1)'_{*}(1,0,-1)'((-1,0,1)',(1,0,-1)')$  or  $(0,-1,1)'_{*}(0,1,-1)'$  and the other columns are zero denoted by  $-z_{*}$  like

cnot	cu by	<b>~</b> ],	IKC		
1	-1	0]	[1	-1	01
-1	1	Ο,	0	0	0
0	0	0	l–1	1	01
0	0	0]	[-1	1	01
1	-1	0,	1	-1	0
-1	1	0	Lo	0	0]
-1	1	0]	[0]	0	01
0	0	0,	-1	1	0.
1	-1	0	l1	-1	0]

Also, we can write all elements of **B**as one-dimensional column vectoras follows:

 $z_j = (z_1, \dots, z_n)', j = 1, \dots, k$  and  $z_t = 1$  or -1 or 0 such that

$$\begin{split} & \lim_{i \neq 1} \|f t = 1, 2, \dots, \frac{n}{2} \\ & \|f t = 1, 2, \dots, \frac{n}{2} \\ & = \begin{cases} 1 \ if \ z_{t+\frac{n}{3}} + z_{t+\frac{2n}{3}} = -1 \ and \ \sum_{\substack{i=1 \\ i \neq t}}^{\frac{n}{3}} z_i = -1 \\ & -1 \ if \ z_{t+\frac{n}{3}} + z_{t+\frac{2n}{3}} = 1 \ and \ \sum_{\substack{i=1 \\ i \neq t}}^{\frac{n}{3}} z_i = 1 \\ & 0 \ if \ z_{t+\frac{n}{3}} + z_{t+\frac{2n}{3}} = 0 \ and \ \sum_{\substack{i=1 \\ i \neq t}}^{\frac{n}{3}} z_i = 0 \\ & \text{If } t = \frac{n}{2} + 1, \frac{n}{2} + 2, \dots, \frac{2n}{2} \\ & \begin{cases} 1 \ if \ z_{t-\frac{n}{3}} + z_{t+\frac{n}{3}} = -1 \ and \ \sum_{\substack{i=1 \\ i \neq t}}^{\frac{2n}{3}} z_i = -1 \\ & \frac{1 \ if \ z_{t-\frac{n}{3}} + z_{t+\frac{n}{3}} = -1 \ and \ \sum_{\substack{i=1 \\ i \neq t}}^{\frac{2n}{3}} z_i = -1 \\ & \\ -1 \ if \ z_{t-\frac{n}{3}} + z_{t+\frac{n}{3}} = 1 \ and \ \sum_{\substack{i=1 \\ i \neq t}}^{\frac{2n}{3}} z_i = 1 \\ & 0 \ if \ z_{t-\frac{n}{3}} + z_{t+\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=1 \\ i \neq t}}^{\frac{2n}{3}} z_i = 0 \\ & \text{If } t = \frac{2n}{2} + 1, \frac{2n}{2} + 2, \dots, n \\ & 1 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = -1 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = -1 \\ & -1 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = -1 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = -1 \\ & -1 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 1 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 1 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ \sum_{\substack{i=\frac{2n}{3} + 1 \\ i \neq t}}^{n} z_i = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \ and \ z_{t-\frac{2n}{3}} = 0 \\ & 0 \ if \ z_{t-\frac{2n}{3}} + z_{t-\frac{n}{3}} = 0 \\$$

Volume 3 Issue 12, December 2014 <u>www.ijsr.net</u> Licensed Under Creative Commons Attribution CC BY **Theorem 2 (see [7]).** The number of elements in **B** equal to  $n^2 - 2n$ 

**Remark** 5(see [7]).Given a contingency table  $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ , the entry of the matrix  $\mathbf{A}$  in the column indexed by  $x_1, x_2, \dots, x_n$  respectively and its rows indexed by

$$\sum_{i=1}^{n} x_{i}, \sum_{i=\frac{n}{2}+1}^{\frac{2n}{3}} x_{i}, \sum_{i=\frac{n}{2}+1}^{n} x_{i}, \sum_{i=\frac{2n}{3}+1}^{n} x_{i}, x_{1} + x_{\frac{n}{3}+1} + x_{\frac{2n}{3}+1}, x_{2} + x_{\frac{n}{3}+2} + x_{\frac{2n}{3}+1}, x_{3} + x_{3} +$$

respectively. The entry in the column indexed by  $x_{i}$  in the matrix **A** will be equal to one, if  $x_{i}$  a pears in the index of its row, and otherwise it will be zero. Then

А															
[	1	1	 1	1	0	0		0	0	0	0		0	01	
	0	0	 0	0	1	1		1	1	0	0		0	0	
	0	0	 0	0	0	0		0	0	1	1		1	1	
=	1	0	 0	0	1	0	***	0	0	1	0	***	0	0	
	0	1	 0	0	0	1	***	0	0	0	1		0	0	
	1	1					***			1					
	0	0	 0	1	0	0		0	1	0	0		0	1	n+9

**Theorem 3 (see [7]).**The set  $\mathbf{B} = \{\mathbf{z}_1, \dots, \mathbf{z}_{\frac{n^2 - 2n}{3}}\}$  is a set of moves.

**Corollary 1 (see [7]).**The set **B** of moves in theorem10 is a Markov basis.

**Corollary 2** (see [7]). The toric ideal  $I_A$  for  $\frac{n^2 - 2n}{2} \times 3 \times \frac{n}{2}$  - contingency tables are  $I_A = \langle P_{i+1}P_{j+k} - P_{j+1}P_{i+k}; i, j = 1, 2, ..., \frac{n}{2}$  and  $l, k = 0, \frac{n}{2}, \frac{2n}{2}$ , such that i < j and  $l < k > \subset \mathcal{C}[P_1, P_2, ..., P_n]$ .

Remark 6 (see [7]). Now, we will construct a connected graph by using the elements of **B**. Let  $z_m$  be an element of Busch that  $\mathbf{z}_m = \mathbf{x}_m - \mathbf{x}_{m-1}$ ,  $m = 1, 2, \dots, \frac{n^2 - 3n}{3}$  is an edge connected  $x_m$  and  $x_{m-1}$ , ..., and  $z_{\frac{n^2-3n}{3}} = x_0 - x_{\frac{n^2-3n}{3}-1}$  is an edge connect  $\mathbf{x}_0$  an  $\mathbf{x}_{\underline{n}^2-\underline{3}\underline{n}}$ , where  $\mathbf{x}_i \in A^{-1}[t], i = 0, 2, \dots, \frac{n^2-\underline{3}\underline{n}}{\underline{3}} - 1$ . Then and we can connected all  $\frac{n^2 - 3n}{3} \times 3 \times \frac{n}{3}$  - contingency tables with fixed two dimensional marginals by  $\frac{m^2-2m}{2}$  edges by applying moves from  ${\bf B}$  one by one and go from  ${\pmb x}_{\tt 0}$  to  $x_{n^2-3n} = 1$  without causing negative cell frequencies on the way, and also from  $x_{\underline{n^2-3n}-1}$  to  $x_{0}$ , of this type, by forming undirected graph  $G = (R, W, B) = A^{-1}[t]_B$ , where the contingency tables interpreted as vertices and connecting moves are interpreted as edges of a graph,  $I_A$ 

, such that i < j and  $l < k > \subset C[P_1, P_2, \dots, P_n]$ , for all  $g \in H$ .

 $R = \{x_0, x_2, \dots, x_{\frac{n^2 - 3n}{3} - 2}\} \text{ and } W = \{x_1, x_2, \dots, x_{\frac{n^2 - 3n}{3} - 1}\}$ as



In [8] H. H. Abbass and H. S. Mohammed Hussein assumed *n* is a multiple of 6, and *H* is the subgroup  $\{e, r^{\frac{n}{3}}, r^{\frac{2n}{3}}, sr, sr^{\frac{n}{3}+1}, sr^{\frac{2n}{3}+1}\}$  of dihedral group  $D_n$ , where  $r = (123 \dots n)$  and  $s = (2n)(3n-1) \dots (\frac{nn}{2n}+2)$ .

**Theorem 4 (see [7]).**The graph G = (R, W, B) is a connected bipartite graph (up to graph isomorphism).

**Theorem5 (see [8]).** The Markov basis**B** is **H**-invariant. **Corollary3 (see [8]).** The **subgroup** H is the Largest Subgroup of the group  $D_n$  such that the Markov basis**B** is **H**-invariant.

Remark 7 (see [8]). Let  $t = (t_1, t_2, t_3, \dots, t_{\frac{n}{3}+3})^{t}$ ,  $x_i \in A^{-1}[t]$  and  $g \in H$ . Then  $gx_i \in A^{-1}[gt]$  where  $gt = (gt_1, gt_2, gt_3, \dots, gt_{\frac{n}{3}+3})^{t}$ ,  $A^{-1}[gt] = \{x \in \mathbb{N}^n : Ax = gt\}$ . So, we have six types of gt-fibers  $A^{-1}[t]$ ,  $A^{-1}[r^{\frac{n}{3}}t]$ ,  $A^{-1}[r^{\frac{2n}{3}}t]$ ,  $A^{-1}[srt]$ ,  $A^{-1}[sr^{\frac{n}{3}+1}t]$  and  $A^{-1}[sr^{\frac{2n}{3}+1}t]$ . Theorem 6 (see [8]). If  $g \in H$ , then **B** is a Markov basis for

 $\frac{n^2 - \exists n}{\exists} \text{ contingency tables} gx_0, gx_1, \dots, gx_{\frac{n^2 - \exists n}{3} - 1} \text{ in } A^{-1}[gt].$ Corollary 4 (see [8]). The toric ideal for  $\frac{n^2 - \exists n}{3} \times 3 \times \frac{n}{3}$  - contingency table in  $A^{-1}[gt]$  is

$$= < P_{g(i+l)}P_{g(j+k)} - P_{g(j+l)}P_{g(i+k)}: i, j = 1, 2, \dots, \frac{n}{3} \text{ and } l, k = 0, \frac{n}{3}, \frac{2n}{3}$$

#### 3. The Main Results

Let **n** is multiple of 6,  $gx_j \in A^{-1}[gt], j = 0, \dots, \frac{n^2 - 3n}{3} - 1$  and  $g \in H$  be representative elements of the set of  $3 \times \frac{n}{3}$  -contingency

#### Volume 3 Issue 12, December 2014 www.ijsr.net

Paper ID: OCT14464

Licensed Under Creative Commons Attribution CC BY

545

International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358

tables. Then we write g as  $n \times n$  permutation matrix  $T_g = \{P_{ij}\} = \{\delta_{i}, g(i)\}$ , where  $\delta$  is the Kronecker's delta such that  $T_{g_1g_2} = T_{g_1}T_{g_2}$  for  $g_1, g_2 \in H$ , and  $T_g^{-1} = T_g'$ . The identity matrix of the order n denoted by  $E_n$  for the unit element e.

Now, we consider a left action of dihedral group  $D_n \cdot n = |I|$ , on  $A^{-1}[t]$  the set of  $\frac{n^2 - 2n}{2} \times 3 \times \frac{n}{2}$  - contingency tables, and the action of dihedral group  $D_n$  on the set of Markov basis **B**.

**Theorem 7.** Let  $x_i, x_i \in A^{-1}[t]$ , if  $g \in H$ . Then  $x_i$ accessible from  $x_j$  by **B** if and only if  $T_g x_i$  accessible from  $T_g \mathbf{x}_j$  by **B**, for all  $i, j = 1, 2, \dots, \frac{n^2 - 3n}{3}$ . **Proof:** If  $x_i$  accessible from  $x_i$  by **B**, then there exists a of moves  $z_1, \dots, z_k \in \mathbf{B}$ sequence and 
$$\begin{split} & \varepsilon_{\mathbf{k}} \in \{-1, 1\}, \mathbf{k} = 1, \dots, \frac{n^2 - 2n}{3}, \qquad \text{such} \qquad \text{that} \\ & x_i = x_j + \sum_{k=1}^{\frac{n^2 - 3n}{3}} \varepsilon_k z_k, x_j + \sum_{k=1}^{\frac{n^2 - 3n}{3}} \varepsilon_k z_k \in A^{-1}[t] \text{ for} 1 \le k \le \frac{n^2 - 3n}{3}. \end{split}$$
(definition 9) Let  $g \in H$ If  $g = e \implies E_n \mathbf{x}_i = E_n \mathbf{x}_j + \sum_{k=1}^{\frac{n^2 - sn}{s}} \varepsilon_k E_n \mathbf{z}_k E_n \mathbf{x}_j$  $+ \sum \varepsilon_k E_n \mathbf{z}_k \in H(A^{-1}[\mathbf{t}])$ for  $1 \le k \le \frac{n^2 - 3n}{2}$ . If  $g = r^{\frac{n}{3}}$  then we write  $r^{\frac{n}{3}}$  as a  $n \times n$  permutation matrix, i.e.  $r^{\frac{n}{3}} = T_{\left(1\frac{n}{3}+1\frac{2n}{3}+1\right)\left(2\frac{n}{3}+2\frac{2n}{3}+2\right)\dots\left(\frac{n2n}{3}n\right)}$ , then  $T_{\left(1,\frac{n}{3}+1,\frac{2n}{3}+1\right)\left(2,\frac{n}{3}+2,\frac{2n}{3}+2\right)\dots\left(\frac{n}{3},\frac{2n}{3},n\right)}x_{i} = T_{\left(1,\frac{n}{3}+1,\frac{2n}{3}+1\right)\left(2,\frac{n}{3}+2,\frac{2n}{3}+2\right)\dots\left(\frac{n}{3},\frac{2n}{3},n\right)}x_{j}$  $+ \sum_{k=1}^{\frac{n}{3}} \varepsilon_k T_{\left(1,\frac{n}{3}+1,\frac{2n}{3}+1\right)\left(2,\frac{n}{3}+2,\frac{2n}{3}+2\right),..,\left(\frac{n}{3},\frac{2n}{3},n\right)} z_k,$  $T_{\left(1\frac{n}{3}+1\frac{2n}{3}+1\right)\left(2\frac{n}{3}+2\frac{2n}{3}+2\right)...\left(\frac{n}{3}\frac{2n}{3}n\right)x_{j}}^{n+1}$  $\sum_{k=1}^{3} \varepsilon_k T_{\left(1 \frac{n}{3}+1 \frac{2n}{3}+1\right)\left(2 \frac{n}{3}+2 \frac{2n}{3}+2\right)...\left(\frac{n}{3} \frac{2n}{3} n\right)} z_k$  $\in H(A^{-1}[t])$  for  $1 \le k \le \frac{n^2 - 3n}{2}$ . If  $g = r^{\frac{2n}{3}}$ , then  $r^{\frac{2n}{3}} = T_{\left(1\frac{2n}{3}+1\frac{n}{3}+1\right)\left(2\frac{2n}{3}+2\frac{n}{3}+2\right)\dots\left(\frac{n}{3}n\frac{2n}{3}\right)}$  $T_{\left(1\frac{2n}{3}+1\frac{n}{3}+1\right)\left(2\frac{2n}{3}+2\frac{n}{3}+2\right)\dots\left(\frac{n}{3}n\frac{2n}{3}\right)}\mathbf{x}_{i}} = T_{\left(1\frac{2n}{3}+1\frac{n}{3}+1\right)\left(2\frac{2n}{3}+2\frac{n}{3}+2\right)\dots\left(\frac{n}{3}n\frac{2n}{3}\right)}\mathbf{x}_{i}}$ +  $\sum_{k=1}^{3} \varepsilon_k T_{\left(1 \frac{2n}{3}+1 \frac{n}{3}+1\right)\left(2 \frac{2n}{3}+2 \frac{n}{3}+2\right)\dots\left(\frac{n}{3} n \frac{2n}{3}\right)} \mathbf{z}_{k_i}$  $T_{(1\frac{2n}{3}+1\frac{n}{3}+1)(2\frac{2n}{3}+2\frac{n}{3}+2)\dots(\frac{n}{3}n\frac{2n}{3})}x_{j} +$  $\sum_{k=1}^{\frac{n}{3}} \tilde{\varepsilon}_k T_{\left(1 \frac{2n}{3}+1 \frac{n}{3}+1\right)\left(2 \frac{2n}{3}+2 \frac{n}{3}+2\right)...\left(\frac{n}{3} n \frac{2n}{3}\right)} \mathbf{z}_k$ 

$$\begin{split} & \in H(A^{-1}[t]) \text{ for } 1 \leq k \leq \frac{n^2 - 3n}{2} \\ & \text{ Hg = sr, then } sr = T_{(1\,n)(2\,n-1) \dots (\frac{2n}{2}, \frac{n}{2}+1)}, \\ & T_{(1\,n)(2\,n-1) \dots (\frac{n}{2}, \frac{n}{2}+1)} x_i = T_{(1\,n)(2\,n-1) \dots (\frac{n}{2}, \frac{n}{2}+1)} x_j \\ & \frac{n^2 - an}{s} \\ & + \sum_{k=1}^{n^2} \varepsilon_k T_{(1\,n)(2\,n-1) \dots (\frac{n}{2}, \frac{n}{2}+1)} x_j + \sum_{k=1}^{n^2} \varepsilon_k T_{(1\,n)(2\,n-1) \dots (\frac{n}{2}, \frac{n}{2}+1)} x_k \\ & \\ & T_{(1\,n)(2\,n-1) \dots (\frac{n}{2}, \frac{n}{2}+1)} x_j + \sum_{k=1}^{n^2} \varepsilon_k T_{(1\,n)(2\,n-1) \dots (\frac{n}{2}, \frac{n}{2}+1)} x_k \\ & \\ & \in H(A^{-1}[t]) \text{ for } 1 \leq k \leq \frac{n^2 - 3n}{3}. \\ & \text{ Hg = } sn^{\frac{n}{2}+1} \qquad \text{ then } \\ & sr^{\frac{n}{2}+1} = T_{(1,\frac{2n}{3})(2,\frac{2n}{3}-1) - (\frac{3n}{3}+1)(\frac{2n}{3}+1n)(\frac{2n}{3}+2n-1)} \\ & \dots (\frac{5n5n}{8\,e^{n}+1})^{x_i} = \\ & T_{(1,\frac{n}{3})(2,\frac{2n}{3}-1) - (\frac{nn}{3}+1)(\frac{2n}{3}+1n)(\frac{2n}{3}+2n-1)} \\ & \dots (\frac{8n5n}{8\,e^{n}+1})^{x_i} = \sum_{k=1}^{n^2-2n} \\ & -(\frac{8n5n}{8\,e^{n}+1})^{x_i} + \sum_{k=1}^{2} \varepsilon_k T_{(1,\frac{2n}{3})(2,\frac{2n}{3}-1)} \\ & -(\frac{8n5n}{8\,e^{n}+1})^{x_i} + \frac{n^2 - 2n}{2} \\ & \sum_{k=1}^{n^2-2n} \varepsilon_k T_{(1,\frac{2n}{3})(2,\frac{2n}{3}-1) - (\frac{8n}{3n}+1)(\frac{2n}{3}+2n-1)} \\ & -(\frac{8n5n}{8\,e^{n}+1})^{x_i} + \frac{n^2 - 2n}{2} \\ & \sum_{k=1}^{n^2-2n} \varepsilon_k T_{(1,\frac{2n}{3})(2,\frac{2n}{3}-1) - (\frac{2n}{3n}+1)(\frac{2n}{3}+2n-1)} \\ & -(\frac{8n5n}{8\,e^{n}+1})^{x_i} + \frac{n^2 - 2n}{3} \\ & \sum_{k=1}^{n^2-3n} \varepsilon_k T_{(1,\frac{2n}{3})(2,\frac{2n}{3}-1) - (\frac{2n}{3n}+1)(\frac{2n}{3}+2n-1)} \\ & (\frac{8n5n}{3}+1)^{x_i} + \frac{n^2 - 3n}{3} \\ & \text{If} \\ g = sr^{\frac{2n}{3}+1}, \text{then } sr^{\frac{2n}{3}+1} = T_{(1,\frac{2n}{3})(2,\frac{n}{3}-1) - (\frac{n}{3}+1n)(\frac{n}{3}+2n-1) - (\frac{2n}{3n}\frac{2n}{3}+1)^{x_i} \\ & T_{(1,\frac{n}{3})(2,\frac{n}{3}-1) - (\frac{n}{3}+1,n)(\frac{n}{3}+2n-1) - (\frac{2n}{3}\frac{2n}{3}+1)^{x_i} \\ & T_{(1,\frac{n}{3})(2,\frac{n}{3}-1) - (\frac{n}{3}+1,n)(\frac{n}{3}+2n-1) - (\frac{2n}{3}\frac{2n}{3}+1)^{x_i} \\ & T_{(1,\frac{n}{3})(2,\frac{n}{3}-1) - (\frac{n}{3}+1,n)(\frac{n}{3}+2n-1) - (\frac{2n}{3}\frac{2n}{3}+1)^{x_i} \\ & T_{(1,\frac$$

Conversely,

Volume 3 Issue 12, December 2014

<u>www.ijsr.net</u> Licensed Under Creative Commons Attribution CC BY

If  $T_{\alpha} x_i$  accessible from  $T_{\alpha} x_i$  by  $H(\mathbf{B})$ , then there exists a sequence of moves  $gz_1, ..., gz_s \in H(\mathbf{B}), g \in H$ , and  $\varepsilon_k \in \{-1, 1\}, k = 1, ..., \frac{n^2 - 3n}{3}$ , such that  $T_{g}\boldsymbol{x}_{i} = T_{g}\boldsymbol{x}_{j} + \sum_{i=1}^{3} \varepsilon_{k} T_{g} \boldsymbol{z}_{k},$  $T_{g}x_{j} + \sum_{k=1}^{\frac{n^{2}-3n}{3}} \varepsilon_{k} T_{g}z_{k} \in H(A^{-1}[t]) \text{ for } 1 \le k \le \frac{n^{2}-3n}{3}.$ (definition 9

By multiplying the previous equations by  $T'_{g^*}g \in H$ , we have

$$T'_{g}(T_{g}\boldsymbol{x}_{i}) = T'_{g}(T_{g}\boldsymbol{x}_{j}) + \sum_{k=1}^{\frac{n^{2}-3n}{3}} \varepsilon_{k} T'_{g}(T_{g}\boldsymbol{z}_{k}),$$

$$T'_{g}(T_{g}\boldsymbol{x}_{j}) + \sum_{k=1}^{\frac{n^{2}-3n}{3}} \varepsilon_{k} T'_{g}(T_{g}\boldsymbol{z}_{k}) \in H(A^{-1}[\boldsymbol{t}]) \text{ for } 1 \le k \le \frac{n^{2}-3n}{3}$$
This implies,

This

$$(T'_{g}T_{g})x_{i} = (T'_{g}T_{g})x_{j} + \sum_{k=1}^{\frac{n^{2}-3n}{3}} \varepsilon_{k}(T'_{g}T)_{g}z_{k},$$
  
$$(T'_{g}T_{g})x_{j} + \sum_{k=1}^{\frac{n^{2}-3n}{3}} \varepsilon_{k}(T'_{g}T_{g})z_{k} \in \mathcal{H}(A^{-1}[t]) \text{ for } 1 \le k \le \frac{n^{2}-3n}{3}$$

Hence 
$$E_n \mathbf{x}_i = E_n \mathbf{x}_j + \sum_{k=1}^{n-2n} \varepsilon_k E_n \mathbf{z}_k$$
  
 $E_n \mathbf{x}_j + \sum_{k=1}^{n^2-2n} \varepsilon_k E_n \mathbf{z}_k \in H(A^{-1}[t]) \text{ for } 1 \le k \le \frac{n^2 - 3n}{3}$ 

Therefore,  $\mathbf{x}_i = \mathbf{x}_j + \sum_{k=1}^{\frac{n^2 - 3n}{3}} \varepsilon_k \mathbf{z}_k$  $x_j + \sum_{k=1}^{\frac{n^2 - 3n}{3}} \varepsilon_k x_k \in A^{-1}[t] \text{ for } 1 \le k \le \frac{n^2 - 3n}{3}.$  [since  $E_n x_i = x_i$  and  $E_n z_k = z_k$ Then  $x_i$  accessible from  $x_i$  by **B**.

Remark 8.Now, we will construct a connected graph by using the elements of  $H(\mathbf{B})$ . Let  $g\mathbf{z}_k = T_g \mathbf{z}_k$  be an element of  $H(\mathbf{B})$  $\mathbf{g} \in \mathbf{H}$ such for all that  $g \mathbf{z}_{k} = g \mathbf{x}_{k} - g \mathbf{x}_{k-1} = T_{g} \mathbf{x}_{k} - T_{g} \mathbf{x}_{k-1}$  $\mathbf{x}_k = \mathbf{1}, \mathbf{2}, \dots, \frac{n^2 - 2n}{2}$  is an edge connected  $g\mathbf{x}_k = T_g \mathbf{x}_k$  and  $g\mathbf{x}_{k-1} = T_g \mathbf{x}_{k-1}$  $g z_{\frac{n^2 - 3n}{3}} = g x_0 - g x_{\frac{n^2 - 3n}{3} - 1} = T_g x_0 - T_g x_{\frac{n^2 - 3n}{3} - 1}$  be an edge connected  $T_g x_0$  and  $T_g x_{\frac{n^2 - 3n}{3} - 1}$ , where  $gx_i \in A^{-1}[gt], g \in H \ i = 0, 1, 2, ..., \frac{n^2 - 3n}{n} - 1$ . Then we canconnect all  $\frac{n^2-3n}{3} \times 3 \times \frac{n}{3}$  - contingency tables with fixed two dimensional marginals by  $\frac{n^2-3n}{2}$  edges by applying moves from  $H(\mathbf{B})$  to  $T_g \mathbf{x}_0$  one by one and go from  $T_g \mathbf{x}_0$  to  $T_g x_{n^2-3n} = 1$  without causing negative cell frequencies on the

way, and also from  $T_g \mathbf{x}_{\underline{n^2}-\underline{n}\underline{n}}$  to  $T_g \mathbf{x}_{\underline{0}}$ . This forms

undirected graph as shown in figure 2



Figu

, wh

conr T<sub>o</sub>R

Figure 2. The graph 
$$T_g G = (T_g R, T_g W, H(\mathbf{B})) = \mathbf{A}^{-1}[gt]_{H(\mathbf{B})} = \mathbf{A}^{-1}[gt]_{\mathbf{B}}$$
, where the contingency tables interpreted as vertices and connecting moves are interpreted as edges of a graph,  $T_g R = \{T_g \mathbf{x}_0, T_g \mathbf{x}_2, \dots, T_g \mathbf{x}_{\frac{n^2 - 3n}{3} - 2}\}$  and

$$T_g W = \{T_g x_1, T_g x_2, \dots, T_g x_{\frac{n^2 - 3n}{3} - 1}\}.$$

**Theorem 8.** The graphs  $T_g G = (T_g R, T_g W, H(\mathbf{B}))$  are connectedbipartite graphs (up to graph isomorphism). **Proof**:

Let 
$$T_g x_i, T_g x_j \in A^{-1}[gt]$$
, if

 $0 \le i \le j \le \frac{n^2 - 2n}{n} - 1, i \ne j_i$  by remark 8there exists a path  $< T_g x_i, T_g z_{i+1}, T_g x_{i+1}, T_g z_{i+2}, \dots, T_{\underline{n}} x_{j-1}$ 

$$T_{r\overline{s}} \mathbf{z}_{j}, T_{r\overline{s}} \mathbf{x}_{j} > , \text{ and if } \mathbf{0} \le j \le i \le \frac{n^{2} - 2n}{2} - 1, i \ne j, \text{ by}$$
  
remark 8there exists a path  $< T_{g} \mathbf{x}_{j}, T_{g} \mathbf{z}_{j+1}, T_{g} \mathbf{x}_{j+1}, T_{g} \mathbf{z}_{j+2}, \dots, T_{g} \mathbf{x}_{i-1},$ 

 $T_{\sigma}z_{i}, T_{\sigma}x_{i} >$ , and that implies there exists a path between every pair of distinct vertices  $T_{\sigma}x_i, T_{\sigma}x_i \in A^{-1}[t]$  of the graph, by (definition 8), *G* is a connected graph.

Now, we prove the graph  $G = (T_g R_s T_g W_s H(B))$  is a bipartite graph.

Let  $T_g \mathbf{x}_i, T_g \mathbf{x}_{i+1}, \dots, T_g \mathbf{x}_{j-1}, T_g \mathbf{x}_j, T_g \mathbf{x}_{j+1} = T_g \mathbf{x}_i$  be a cycle in G. Suppose  $T_{\sigma}x_i \in T_{\sigma}R$ . Then  $T_{\sigma}x_{i+1} \in T_{\sigma}W$ , since the  $edgeT_{\sigma}\mathbf{z}_{i+1} = T_{\sigma}\mathbf{x}_{i+1} - T_{\sigma}\mathbf{x}_i \in H(\mathbf{B}),$ then  $T_{\sigma} \mathbf{x}_{i+2} \in T_{\sigma} R_{\sigma}$  since the edge

 $T_g \mathbf{z}_{i+2} = T_g \mathbf{x}_{i+2} - T_g \mathbf{x}_{i+1} \in H(\mathbf{B})$ . Continuing in this way, we see that if k is an odd, then  $T_{\sigma} x_k \in T_{\sigma} W$ , and if k is even, then  $T_g x_k \in T_g R$ . Since  $T_g x_{i+1} = T_g x_i \in T_g R$ , it implies that j + 1 is even and thus the cycle is of even length. By theorem 1, then the graph  $G = (T_g R_s T_g W_s H(\mathbf{B}))$  is a bipartite graph.

#### 4. Genomics and Phylogenetic

In this section, we describe some of the basic biological facts needed to understand phylogenetic models and then delve into the practical side of the algebraic statistics of these models. The basic genetic information of an organism is (almost always) carried in the form of DNA, a double

#### Volume 3 Issue 12, December 2014 www.ijsr.net

helix consisting of two complementary B polymers bound together. The DNA molecules in a genome are typically represented as a number of frequencies of letters from the four letters alphabet= {A, C, G, T}. These letters correspond to the bases in the double helix that is the nucleotides Adenine, Cytosine, guanine and Thymine. The four nucleotides that form DNA come in two types: the purines (A and G) and the pyrimidine's (C and T). The two strands of the double helix are joined together via the base pairings A to T (via 2 hydrogen bonds) and C to G (via 3 hydrogen bonds). Since each cell typically contains a copy of the DNA of the organism, DNA copying occurs frequently. Several types of errors are possible during the replication of DNA. Single bases can mutate, or large pieces of DNA can separate and become reattached, possibly at another position, possibly in the opposite direction, these are just some of the events that occur over the course of evolution (C. Semple, M [5] and J. Felsenstein[9]).

# **5.** A New Model of Genetic Algorithm Using the Action of Largest Subgroup of $D_n$ for Invariance Markov Basis

In this section, we construct a new model of genetic algorithm that permutes the pieces of nucleotides in aligned DNA sequences using the actions of largest subgroup H of  $D_n$  for invariance Markov basis and toric ideals. Now, we describe our model in the following steps.

<u>Step (1)</u>:Suppose we have l-taxons of DNA sequences each taxion of length L such as

Taxon1: A G C T A A CG G T A T ···· Taxon2: C G A T C T G A C C T T ····

TaxonI: A C G T C A C G T A G C \*\*\*

Now, we define a pattern  $i = i_1, i_2, ..., i_m$  to be the sequence of characters. We look at a single site (column) of our sequence data. In the sequences above, we can look at the first site in the sequences and see the pattern "AC . . . A". A pattern frequency  $x_i$  is that *i* appears in our set of sequence data , and we denote to the number of frequencies by *n* where *n* is a multiple of 6.

<u>Step (2)</u>: We can input pattern frequency  $x_i$  of above sequences in  $\Im \times \frac{\pi}{2}$  -contingency table as follows :

<i>x</i> <sub>1</sub>	<i>x</i> <sub>2</sub>	 <u>x n</u> 3	$\sum_{i=1}^{\frac{n}{3}} x_i$
x <u>n</u> 3+1	$\frac{x_{\frac{n}{3}+2}}{x_{\frac{n}{3}+2}}$	 x 2n 3	$\sum_{i=\frac{n}{3}+1}^{\frac{2n}{3}} x_i$
x <u>2n</u> 3+1	$\frac{x_{\frac{2n}{3}+2}}{x_{\frac{2n}{3}+2}}$	 xn	$\sum_{i=\frac{2n}{3}+1}^{n} x_{i}$
$x_1 + x_{\frac{n}{3}+1} + x_{\frac{2n}{3}+1}$	$ \begin{array}{c} x_2 \\ + x_{\frac{n}{3}+2} \\ + x_{\frac{2n}{3}+2} \end{array} $	 $\frac{x_{\frac{n}{3}} + x_{\frac{2n}{3}}}{x_n}$	$ \mathbf{x}  = \sum_{i \in I} x_i$

Where  $|\mathbf{x}| = \sum_{i \in I} x_i = L$  is the length of sequences (the sample size), and

 $x_1$  is the frequency of the first pattern .

 $x_2$  is the frequency of the second pattern.

 $x_{\frac{m}{2}}$  is the frequency of the  $\frac{m}{2}$  pattern.

 $x_{\frac{n}{2}+1}^{n}$  is the frequency of the  $\frac{n}{4} + 1$  pattern.  $x_{\frac{n}{2}+2}^{n}$  is the frequency of the  $\frac{n}{4} + 2$  pattern.

Ē

 $x_{m}$  is the frequency of the m pattern.

**Step** (3):Represent the contingency table  $\mathbf{x} = \{x_i\}_{i \in I} \in \mathbb{N}^n$  as a *n*-dimensional column vector of nonnegative integers  $\mathbf{x} = (x_1, x_2, \dots, x_n)^t$ , Where' denotes the transpose of a vector or matrix , then  $\mathbf{x}$  is a *t*-fiber (i.e)  $\mathbf{x} \in \mathbb{A}^{-1}[t]$  where  $\mathbb{A}^{-1}[t] = \{\mathbf{x} \in \mathbb{N}^n : \mathbb{A}\mathbf{x} = t\}$ . **Step(4)**: From remark 5, A is  $\frac{n+9}{n} \times n$  matrix and

А													
	<b>[1</b>	1	 1	1	0	0		0	0	0	0	 0	0]
	0	0	 0	0	1	1		1	1	0	0	 0	0
	0	0	 0	0	0	0		0	0	1	1	 1	1
=	1	0	 0	0	1	0		0	0	1	0	 0	0
	0	1	 0	0	0	1		0	0	0	1	 0	0
	1:				10	1	***		1	1	1		- 1
	LO	0	 0	1	0	0		0	1	0	0	 0	1 _ <u>n+9</u> ×n
													3

Where the columns of the matrix  $\mathbb{A}$  index by the elements of the column vector  $\boldsymbol{x}$ .

Step (5): We can find the Markov basis from remark 4.

Step (6):  $A:\mathbb{Z}^{2\times \binom{n}{2}} \to \mathbb{Z}^{2+\binom{n}{2}}$  is a linear transformation,  $t \in \mathbb{Z}^{2+\binom{n}{2}}$ , and  $A^{-1}[t]$  be the set of t-fibers, and  $\mathbf{B} \subset ker_{\mathbb{Z}}(A)$ . Use remark 6 to find the bipartite graph  $G = (R, W, \mathbf{B}) = A^{-1}[t]_{\mathbf{B}}$ .

**<u>Step (7)</u>**: We can find the toric ideals by using corollary 2 for each contingency tables.

Step (8):Find
$$T_g \mathbf{B} = \{T_g \mathbf{z}_1, T_g \mathbf{z}_2, \dots, T_g \mathbf{z}_{\frac{n^2 - 3n}{3}}\} = \mathbf{B}$$
 for all

 $g \in H_s$  where  $T_g$  is a permutation matrix of g.

$$\frac{\text{Step}}{\{T_g \mathbf{x}_0, T_g \mathbf{x}_1, \dots, T_g \mathbf{x}_{\underline{n}^2 - 3n}_{-1}\}} \subseteq \mathbb{A}^{-1}[gt] (gt \text{-fibers}) \text{ for all}$$

$$\begin{split} &g \in H, \text{where} \\ &r^{\frac{n}{2}}t = \left(t_{3}, t_{1}, t_{2}, t_{4}, \dots, t_{\frac{n}{3}+3}\right)', r^{\frac{2n}{3}}t = \left(t_{2}, t_{3}, t_{1}, t_{4}, \dots, t_{\frac{n}{3}+3}\right) \end{split}$$

International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064

Impact Factor (2012): 3.358

$$sr^{\frac{n}{3}+1}t = (t_2, t_1, t_3, t_{\frac{n}{3}+3}, t_{\frac{n}{3}+2}, \dots, t_4)$$
and  
$$sr^{\frac{2n}{3}+1}t = (t_1, t_3, t_2, t_{\frac{n}{3}+3}, t_{\frac{n}{3}+2}, \dots, t_4).$$

51

-- (+ + + +

Step (10):Use remark 8 and theorem 8 to find the graphs  $T_g G = A^{-1}[gt]_B$  for all  $g \in H$ .

Step (11):Use corollary 4 to find the toric ideal  $I_A = \langle P_{g(i+l)}P_{g(j+k)} - P_{g(j+l)}P_{g(i+k)}; i, j = 1, 2, ..., \frac{n}{n}$  and l, k =, such that i < j and l < k > for all  $g \in H$ .

<u>Step (12)</u>: Use the  $\frac{n^2 - 3n}{3} \times 3 \times \frac{n}{3}$  -contingency tables (*t*fibers) in step (6), and  $\frac{n^2 - 3n}{3} \times 3 \times \frac{n}{3}$  -contingency tables (gt-fibers) in step (9) for all  $g \in H$  to find the permutation of nucleotides in aligned DNA sequences.

Example 1:Suppose we have the following three aligned **DNA** sequences

Taxon 1: A G C T G A G G G C T G G A Taxon 2: A A T C T T A A A T C T T T Taxon 3: T C A G AT C C C A G A A T

Step (1): There are three taxons of above DNA sequences with  $|\mathbf{x}| = \sum_{i \in I} x_i = L = 14$  and nine patternsAAT, GAC, CTA, TCG, GTA, ATT with frequencies 1.4.2.2.3 and 2 respectively where n = 6.

<u>Step (2)</u>:Now, we input the patterns frequency  $x_i$  of above sequences in  $3 \times 2$ -contingency table as follows:

1	4	5
2	2	4
3	2	5
6	8	14

Then the table of marginal and conditional probability is:

1	4	5
14	14	14
2	2	4
14	14	14
3	2	5
14	14	14
6	8	1
14	14	

Step (3):Represent the contingency table  $\mathbf{x} = \{\mathbf{x}_i\}_{i \in I} \in \mathbb{N}^6$  as a 6- dimensional column vector of nonnegative integers  $x = (1,4,2,2,3,2)^{t}$ , then x isa t-fiber

(i.e. 
$$x \in A^{-1}[t]$$
, where  $A^{-1}[t] = \{x \in \mathbb{N}^6 : A \begin{vmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{vmatrix} = \begin{bmatrix} 5 \\ 5 \\ 5 \\ 6 \\ 8 \end{bmatrix}$ .)

Step(4): A is  $5 \times 6$  matrix and r1 1 0 0 0 01 0 0 1 1 0 0  $A = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$ , and 1 0 0 1

$$\begin{array}{c} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ \end{array} \right]_{5\times6} \begin{bmatrix} 1 \\ 4 \\ 2 \\ 3 \\ 3 \\ 2 \\ 3 \\ 3 \\ 5\times1 \end{bmatrix}_{6\times1}$$

Where the columns of the matrix A index by the elements of the column vector x, and  $t_1 = x_1 + x_2 = 5, t_2 = x_3 + x_4 = 4, t_3 = x_5 + x_6 = 6, t_4 = 6$  $x_1 + x_3 + x_5 = 7$ ,  $t_5 = x_2 + x_4 + x_6$ 

Step (5): We can find the Markov basis from remark 4. Then the number of moves is  $\frac{n^2 - 3n}{3} = \frac{6^2 - 3 \times 6}{3} = 6$  elements in the set



Step (6): The connected graph  $A^{-1}[t]_{\mathbf{R}} = G = (\mathbf{R}, \mathbf{W}, \mathbf{B})$ with

 $\frac{n^2 - \exists n}{dt} = 6$  (*t*- fibres)  $3 \times 2$ - contingency tables as vertices of it.



Figure 3 : The graph  $G = (R, W, B) = A^{-1}[t]_{R}$ , where the contingency tables interpreted as vertices and connecting Markov basis are interpreted as the edges of a graph,  $R = \{x_0, x_2, x_4\}$  and  $W = \{x_1, x_3, x_5\}$ , where



Volume 3 Issue 12, December 2014 www.ijsr.net Licensed Under Creative Commons Attribution CC BY

549



**Step (7)**: We can find the toric ideals by using corollary 2 for each contingency tables. Then the toric ideals that correspond the Markov basis that shown in the previous figure is:  $I_A = \langle P_1 P_4 - P_2 P_3 , P_1 P_6 - P_2 P_5 , P_3 P_6 - P_4 P_5 \rangle$ .

Step (8):we find the  
set 
$$T_r^2 \mathbf{B} = \{T_r^2 \mathbf{z}_1, T_r^2 \mathbf{z}_2, T_r^2 \mathbf{z}_3, T_r^2 \mathbf{z}_4, T_r^2 \mathbf{z}_5, T_r^2 \mathbf{z}_6\}$$
  
 $= \{\mathbf{z}_3, \mathbf{z}_4, \mathbf{z}_5, \mathbf{z}_6, \mathbf{z}_1, \mathbf{z}_2\} = \mathbf{B}.$   
And we find the

$$setT_{r}^{*}B = \{T_{r}^{*}z_{1}, T_{r}^{*}z_{2}, T_{r}^{*}z_{3}, T_{r}^{*}z_{4}, T_{r}^{*}z_{5}, T_{r}^{*}z_{6}\}$$

$$= \{z_{5}, z_{6}, z_{1}, z_{2}, z_{3}, z_{4}\} = B. \text{ And we find the}$$

$$setT_{sr}B = \{T_{sr}z_{1}, T_{sr}z_{2}, T_{sr}z_{3}, T_{sr}z_{4}, T_{sr}z_{5}, T_{sr}z_{6}\}$$

$$= \{z_{3}, z_{2}, z_{1}, z_{6}, z_{5}, z_{4}\} = B.$$

And we find the set  $T_{sr^3} \mathbf{B} = \{T_{sr^3} \mathbf{z}_1, T_{sr^3} \mathbf{z}_2, T_{sr^3} \mathbf{z}_2, T_{sr^3} \mathbf{z}_4, T_{sr^3} \mathbf{z}_5, T_{sr^3} \mathbf{z}_6\}$   $= \{\mathbf{z}_1, \mathbf{z}_6, \mathbf{z}_5, \mathbf{z}_4, \mathbf{z}_2, \mathbf{z}_2\} = \mathbf{B}$ . And we find the set  $T_{sr^5} \mathbf{B} = \{T_{sr^5} \mathbf{z}_1, T_{sr^5} \mathbf{z}_2, T_{sr^5} \mathbf{z}_2, T_{sr^5} \mathbf{z}_4, T_{sr^5} \mathbf{z}_5, T_{sr^5} \mathbf{z}_6\}$  $= \{\mathbf{z}_5, \mathbf{z}_4, \mathbf{z}_9, \mathbf{z}_2, \mathbf{z}_1, \mathbf{z}_6\} = \mathbf{B}$ .

**Step (9):** We find the set  $\{T_r^2 x_0, T_r^2 x_1, T_r^2 x_2, T_r^2 x_3, T_r^2 x_4, T_r^2 x_5, T_r^2 x_6, t_r^2 t_r^2$ 



And we find the set  $\{T_{r^4}x_0, T_{r^4}x_1, T_{r^4}x_2, T_{r^4}x_2, T_{r^4}x_4, T_{r^4}x_5, T_{r^4}x_6, \} \subseteq A^{-1}[r^4t] (r^4t \text{-fibers}), \text{ where}$  $r^4t = (4, 6, 5, 7, 8)^t$ . We have.

$$T_{r^{4}} \mathbf{x_{0}} = \begin{bmatrix} 2 & 2 & 4 \\ 3 & 2 & 5 \\ 1 & 4 & 5 \\ 6 & 8 & 14 \end{bmatrix}, T_{r^{4}} \mathbf{x_{1}} = \begin{bmatrix} 1 & 3 & 4 \\ 3 & 2 & 5 \\ 2 & 3 & 5 \\ 6 & 8 & 14 \end{bmatrix}$$

$$T_{r^4} \mathbf{x_2} = \begin{bmatrix} 1 & 3 & 4 \\ 2 & 3 & 5 \\ 3 & 2 & 5 \\ 6 & 8 & 14 \end{bmatrix}, T_{r^4} \mathbf{x_3}$$

set

 $\frac{2}{4}$   $\frac{4}{5}$ 

5

3 2

6

8 14

And we find the  $\{T_{sr} x_0, T_{sr} x_1, T_{sr} x_2, T_{sr} x_2, T_{sr} x_4, T_{sr} x_5, T_{sr} x_6\} \subseteq \mathbb{A}^{-1}[srt]$ 

(srt-fibers), where srt = (6, 4, 5, 8, 7)'. We have.

	2	3	5		2	3	5
	2	2	4		3	1	4
	4	1	5		3	2	5
$T_{sr}\mathbf{x_0} =$	8	6	14	$T_{sr}\mathbf{x_1} =$	8	6	14
[	3	2	5		4	1	5
[	3	1	4		2	2	4
[	2	3	5		2	3	5
$T_{sr}\mathbf{x}_2 =$	8	6	14	, $T_{sr}\mathbf{x_3} =$	8	6	14
			_				
	4	1	5		3	2	5
	1	3	4		1	3	4
	3	2	5		4	1	5
$T_{sr}\mathbf{x_4} =$	8	6	14	$, T_{sr}\mathbf{x_5} =$	8	6	14

And we find the set{ $T_{sr} \ge x_0, T_{sr} \ge x_1, T_{sr} \ge x_2, T_{sr} \ge x_2, T_{sr} \ge x_4, T_{sr} \ge x_5, T_{sr} \ge x_6$ }  $\subseteq A^{-1}[sr^2t](sr^2t$ -fibers), where  $sr^2t = (4, 5, 6, 8, 7)^t$ . We have.



And we find the set  $\{T_{sr} \cdot x_0, T_{sr} \cdot x_1, T_{sr} \cdot x_2, T_{sr} \cdot x_2, T_{sr} \cdot x_4, T_{sr} \cdot x_5, T_{sr} \cdot x_6\} \subseteq A^{-1}[sr^5t](sr^5t\text{-fibers}), \text{ where } sr^5t = (5, 6, 4, 8, 7)', \text{ and }$ 



	4	1	5		3	2	5
-	1	3	4	_	1	3	4
$T_{sr}^{5}\mathbf{x_4} =$	8	6	14	<i>T<sub>sr</sub></i> ₅ <b>x</b> 5 =	8	6	14

Step (10): Use remark 8 and theorem 8 to find the graph  $T_{r^2}G = A^{-1}[r^2t]_B$ .



where the contingency tables interpreted as vertices and

connecting moves are interpreted as edges of a graph,  $T_{r^2}R = \{r^2 x_0, r^2 x_2, r^2 x_4\}$  and  $T_{r^2}W = \{r^2 x_1, r^2 x_2, r^2 x_5\}$ . And find the graph  $T_{r^4}G = A^{-1}[r^4t]_B$ .



Figure 5: The graph  $T_{r^4}G = (T_{r^4}R, T_{r^4}W, H(\mathbf{B})) = \mathbf{A}^{-1}[r^4t]_{H(\mathbf{B})} = \mathbf{A}^{-1}[r^4t]_{\mathbf{B}}$ , where the contingency tables interpreted as vertices and connecting moves are interpreted as edges of a graph,  $T_{r^4}R = \{r^4x_0, r^4x_2, r^4x_4\}$  and  $T_{r^4}W = \{r^4x_1, r^4x_2, r^4x_5\}$ . And find the graph  $T_{\underline{sr}}G = \mathbf{A}^{-1}[\mathbf{srt}]_{\mathbf{B}}$ .



Figure 6: The graph  $T_{sr}G = (T_{sr}R, T_{sr}W, H(\mathbf{B})) = A^{-1}[srt]_{H(\mathbf{B})} = A^{-1}[srt]_{\mathbf{B}}$ , where the contingency tables interpreted as vertices and connecting moves are interpreted as edges of a graph,  $T_{sr}R = \{srx_0, srx_2, srx_4\}$  and  $T_{sr}W = \{srx_1, srx_2, srx_5\}$ . And find the graph  $T_{sr}^{3}G = A^{-1}[sr^{2}t]_{\mathbf{B}}$ .

And find the graph  $T_{sr^3}G = A^{-1}[sr^3t]_{\mathbf{B}}$ .





Volume 3 Issue 12, December 2014 <u>www.ijsr.net</u> Licensed Under Creative Commons Attribution CC BY 
$$\begin{split} T_{sr^3} R &= \{sr^3 x_0, sr^3 x_2, sr^3 x_4\} \text{ and } \\ T_{sr^3} W &= \{sr^3 x_1, sr^3 x_3, sr^3 x_5\}. \end{split}$$

And find the graph  $T_{sr^5}G = A^{-1}[sr^5t]_B$ .



Figure 8: The graph  $T_{sr^5}G = \left(T_{sr^5}R, T_{sr^5}W, H(\mathbf{B})\right) = \mathbf{A}^{-1}[sr^5t]_{H(\mathbf{B})} = \mathbf{A}^{-1}[sr^5t]_{\mathbf{B}} T_{r^2}x_1 =$ 

, where the contingency tables interpreted as vertices and connecting moves are interpreted as edges of a graph,  $T_{sr^5}R = \{sr^5x_0, sr^5x_2, sr^5x_4\}$ and  $T_{sr^5}W = \{sr^5x_1, sr^5x_3, sr^5x_5\}.$ 

Step (11):Use corollary 4 to find the toric ideal  $I_A = < P_g(i+i) P_g(j+i) -$ 

 $P_{g(j+l)}P_{g(i+k)}$ :  $i, j = 1, 2, ..., \frac{n}{2}$  and  $l, k = 0, \frac{n}{2}, \frac{2n}{2}$ , such that i < j and  $l < k > = < P_1P_4 - P_2P_3$ ,  $P_1P_6 - P_2P_5$ ,  $P_3P_6 - P_4P_5$ 

**Step(12):**Use the  $6 \times 3 \times 2$  -contingency tables (*t*-fibers) in step (6),  $6 \times 3 \times 2$  -contingency tables (*gt*-fibers) in step (9) to find the permutation of nucleotides in aligned DNA sequences.

Then the change in the type of DNA sequences under the Markov basis.

Be as Figure 3, Where  $x_0 =$ 

```
① ④ ② ② ③ ②
Taxon 1:A G C T G A G G G C T G G A
Taxon 2:A A T C T T A A A T C T T T
Taxon 3:T C A G A T C C C A G A A T
```

 $x_1 =$ 

-1													
2	3	$\overline{0}$	3	3	2	þ							
Taxon 1: 🛦	G	С	Т	G	A	A	G	G	T	Т	G	G	A
Taxon 2: 🛦	<u></u>	Т	С	Т	Т	A	A	A	С	С	Т	Т	Т
Taxon 3: T	С	A	G	A	Т	Т	С	С	G	G	A	A	T
$x_2 =$													
3	2	$\overline{0}$	3	2	3	þ							
Taxon 1: 🛦	G	С	Т	G	A	A	A	G	T	T	G	A	A
Taxon 2: Å	A	Т	С	T	T	A	A	A	С	С	T	T	T
Taxon 3: T	С	A	G	A	T	T	T	C	G	G	A	T	T
x <sub>2</sub> =													
3	2	2	2	1	$\overline{}$	)							
Taxon 1: 🛦	G	С	Т	G	A	A	A	G	С	T.	<u>A</u> .	<u>A</u> .	A
Taxon 2: 🛦	A	Т	С	T	T	A	A	A	T	С	T	T	T
Taxon 3: T	С	A	G	A	T	T	T	С	A	G	T	Т	T

 $x_4 =$ 



And the change in the type of DNA sequences under the action of  $r^2$  on the set of Markov basis be as Figure 4 where  $T_{r^2}x_0 =$ 





#### International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064 Impact Factor (2012): 3.358

0	<u> </u>	9	3	<u> </u>	ন								
T 1. A	్లా	్రా	<u>چ</u>	2	~	~	~	c	Ŧ	<b>T</b> P	~	~	л
Taxon 1: A	ų,	L	L	Un	<u> </u>	U.	<b>U</b> m	L	L	L	U.	Un	æ
Taxon 2: A	A	Т	С	Т	Т	A	A	T	С	С	Т	Т	Т
Taxon 3: T	C	<u>Å</u>	G	A	T	С	С	A	G	G	A	A	T
$T_{r^4}x_3 =$													
2	2	$\overline{1}$	4	3	2	þ							
Taxon 1: A	G	C	T	G	Ā	A	G	T	T	T	G	G	A
Taxon 2: 🛦	A	Т	C	Т	Т	A	A	С	С	С	Т	Т	Т
Taxon 3: T	С	A	G	A	Т	T	C	G	G	G	A	A	Т
$T_{r^{4}}x_{4} =$													
3		1	4	2	3	þ							
Taxon 1: A	G	С	T	G	A	A	A	T	T	T	G	A	A
Taxon 2: 🛦	A	Т	С	Т	Т	A	A	С	С	С	Т	Т	Т
Taxon 3: T	С	A	G	A	Т	T	Т	G	G	G	A	T	T
$T_{r^4}x_5 =$													
3	1	2	3	1	4								
Taxon 1: A	G	C	T	G	Å	A	A	С	T	T	A	A	A
Town 2 A	A	Т	C	Т	Т	4	А	Т	C	C	Т	т	т
	701		<b>1</b>			2020	ALLIN,	-		_	_	-	-

And the change in the type of DNA sequences under the action of *sr* on the set of Markov basis be as Figure 6 where  $srx_0 =$ 

													_
2	3	2	(2)	4	1	•							
Taxon 1: A	G	Ĉ	T	G	Ā	A	G	G	С	Т	G	G	G
Taxon 2: 🙏	<u> </u>	Т	С	T	T	A	A	A	T	С	Т	T	Т
Taxon 3: T	С	<u>A</u>	G	A	T	T	С	C	A	G	A	A	A
$srx_1 =$													
2	3	3	n	ര	$\overline{(2)}$	}							
Taxon 1: A	Ğ	č	Ť	Ğ	Ă	A	G	G	с	С	G	G	A
Taxon 2: A	A	Т	C	Т	Т	A	Ā	A	Т	Т	Т	Т	Т
Taxon 3: T	С	A	G	A	Т	Т	С	С	A	A	A	A	Т
$srx_2 =$													
(3)	2	(3)	$\overline{\mathbb{O}}$	2	3	)							
Taxon 1: A	G	č	Ť	Ğ	Ă	Å	A	G	С	С	G	A	A
Taxon 2: A	A	Т	С	Т	Т	A	A	A	Т	Т	Т	Т	Т
Taxon 3: T	С	A	G	A	Т	Т	Т	С	A	A	A	Т	Т
srx <sub>3</sub> =													
(4)	<u>(</u> )	(2)	$\overline{(2)}$	$\overline{(2)}$	3	•							
Taxon 1: Ă	Ğ	č	Ť	Ğ	Ă	<u>A</u>	A	A	С	Т	G	A	A
Taxon 2: A	A	T	С	Т	Т	A	A	A	T	С	Т	Т	Т
Taxon 3: T	С	A	G	A	Т	Т	T	Т	A	G	A	Т	Т
$srx_4 =$													
4		<b>(1)</b>	3	3	2	)							
Taxon 1: A	G	Ĉ	T	G	Ā	A	A	A	T	Т	G	G	A
-													510
Taxon Z: A	A	Т	С	Т	Т	A	A	A	С	С	Т	Т	
Taxon 2: A Taxon 3: T	<u>A</u> C	T A	C G	T A	T T	A T	A T	A T	C G	C G	T A	T A	T
Taxon 2: A Taxon 3: T $srx_5 =$	A C	T A	C G	T A	T	A T	A T	A T	G	C G	T A	T A	T
Taxon 2: A Taxon 3: T $srx_5 =$	A C (2)		с <u>с</u> З	T <u>A</u> (4)	T T	A T	A T	A T	G	G	T A	T A	T
Taxon 2: A Taxon 3: T $srx_5 =$ Taxon 1: A	A C (2) G		с <u>G</u> З Т			A T	A T A	A T G	C G T	C G T	T A G	T A G	T G
Taxon 2: A Taxon 3: T $srx_5 =$ (3) Taxon 1: A Taxon 2: A	A C 2 G A		С <u>G</u> 3 Т С			A T A A	A T A A	A T G A	C G T C	C G T C	T A G T	T A G T	T G T
Taxon 2:A Taxon 3:T $srx_5 =$ Taxon 1:A Taxon 2:A Taxon 3:T			C G 3 T C G				A T A A T	A T G A C	C G T C G	C G T C G	T A G T A	T A G T A	T G T A

And the change in the type of DNA sequences under the action of  $sr^2$  on the set of Markov basis be as Figure 7 where  $sr^2x_n =$ 

(	<u>2)</u> (	2)	( <b>4</b> )	1)(	2)(	3								
Taxon 1:	A	G	С	Т	G	A	A	G	С	С	С	G	A	A
Taxon 2:	A	A	Т	С	T	T	A	A	T	T	T	Т	Т	Т
Taxon 3:	Т	С	A	G	A	T	Т	С	A	A	A	A	Т	Т
$sr^{3}x_{1} =$														
(	3)(	1	3	2)(	2)(	3								
Taxon 1:	Ă	G	č	T	G	Ă	A	A	С	С	Т	G	A	A
Taxon 2:	A	A	Т	С	Т	T	A	A	T	Т	C	Т	Т	Т
Taxon 3:	Т	С	A	G	A	Т	Т	Т	A	A	G	A	Т	Т
$sr^{3}x_{2} =$														
(	3	<u>(1)</u>	$\overline{(2)}$	3	3	$\overline{(2)}$	þ							
Taxon 1:	Ă	Ğ	č	Ť	G	Ă	A	A	С	Т	Т	G	G	A
Taxon 2:	A	A	Т	С	Т	Т	A	A	T	С	С	Т	Т	Т
Taxon 3:	Т	С	A	G	A	Т	Т	Т	A	G	G	A	A	Т
$sr^3x_n =$														
	2	2	2	3	•	1	•							
Taxon 1:	2 A	2 G	@ C	3 T	• •	1 A	) A	G	С	т	т	G	G	G
Taxon 1: Taxon 2:	2 A A	2 G A	2 C T	3) T C	€ G T		A A	G A	C T	T C	T C	G T	G T	G T
Taxon 1: Taxon 2: Taxon 3:	2 A A T	2 G A C	2 C T A	3 T C G	€ G T A	① A T T	) A A T	G A C	C T A	T C G	T C G	G T A	G T A	G T A
Taxon 1: Taxon 2: Taxon 3: $sr^3x_4 =$	2 A A T	2 G A C	2 C T A	3) T C G	G G T A		A A T	G A C	C T A	T C G	T C G	G T A	G T A	G T A
Taxon 1: Taxon 2: Taxon 3: $sr^{3}x_{4} =$		2 G A C 3	2 C T A 30	3 T C G			A A T	G A C	C T A	T C G	T C G	G T A	G T A	G T A
Taxon 1: Taxon 2: Taxon 3: $sr^3x_4 =$ ( Taxon 1:		2 6 4 C 3 6		3 T C G 2 T			A A T	G A C	C T A C	T C G	T C G	G T A G	G T A G	G T A
Taxon 1: Taxon 2: Taxon 3: $sr^3x_4 =$ (Taxon 1: Taxon 2:				3 T C G Z T C			A A T	G A C G A	C T A C T	T C G C T	T C G T C	G T A G T	G T A G T	G T A G T
Taxon 1: Taxon 2: Taxon 3: $sr^3x_4 =$ (Taxon 1: Taxon 1: Taxon 2: Taxon 2: Taxon 3:				3 T C G 2)( T C G			A A T G A C	G A C A C	C T A C T A	T C G T A	T G T C G	G T A G T A	G T A G T A	G T A G T A
Taxon 1: Taxon 2: Taxon 3: $sr^{2}x_{4} =$ ( Taxon 1: Taxon 2: Taxon 2: Taxon 3: $sr^{3}x_{5} =$		2 G A C 3 G A C		3 T C C T C G			A A T G A C	G A C G A C	C T A T A	T C G T A	T G T C G	G T A T A	G T A T A	G T A T A
Taxon 1: Taxon 2: Taxon 3: $sr^3x_4 =$ ( Taxon 1: Taxon 1: Taxon 2: Taxon 2: $sr^3x_5 =$							A A T G A C	G A C C	C T A C T A	T C G T A	T C G T C G	G T A G T A	G T A G T A	G T A T A
Taxon 1: Taxon 2: Taxon 3: $sr^3x_4 =$ ( Taxon 1: Taxon 1: Taxon 2: Taxon 2: Taxon 3: $sr^3x_5 =$ ( Taxon 1:							A A T G A C	G A C G A C	C T A C T A C	T C C T A C	T G T C G	G T A G T A G	G T A G T A G	G T A G T A
Taxon 1: Taxon 2: Taxon 3: $sr^3x_4 =$ ( Taxon 1: Taxon 2: Taxon 2: $sr^3x_5 =$ ( Taxon 1: $sr^3x_5 =$							G A C G A C	G A C G A C	C T A C T A C T	T C C T A C T	T C G T C G	G T A G T A G T	G T A G T A G T	G T A T A T

And the change in the type of DNA sequences under the action of  $sr^5$  on the set of Markov basis be as Figure 8 where



3	2	٩	1	1	3	)							
Taxon 1: Å	G	С	Т	G	A	A	A	G	С	С	С	A	A
Taxon 2: 🛦	<u> </u>	Т	С	Т	Т	A	A	A	T	T	T	T	Т
Taxon 3: T	С	<u>A</u> .	G	A	Т	Т	T	C	A	A	A	Т	T
5													
$sr^{-}x_{5} =$													
$sr^x_5 =$	1	3	2	1	3	)							
$sr^x x_5 =$ (4) Taxon 1: A	1 G	3) C	20 T	1 G	3 A		A	A	С	С	Т	A	A
$\frac{sr^{-}x_{5}}{4}$ Taxon 1: A Taxon 2: A	① G A	30 C T	20 T C	① G T	3 A T	A	AA	AAA	C T	C T	T C	A T	A T

**Remark 9:** We refer to (1), (2), (3), (4), (5) and (6) in example 1 to the frequencies of the patterns in DNA sequences.

### References

- [1] A.Dobra, "Markov bases for decomposable graphical models", Paper presented at Grostat V, New Oeleans, 4-6 September 2001, submitted for publication, 2001.
- [2] Dobra, and S. Sullivant, "A divide-and-conquer algorithm for generating Markov bases of multi-way tables", Paper presented at the workshop "Multidimensional tables" statistics, combinatorial optimization, and Groebner bases", University of California at Davis, submitted for publication, 15-16 February 2002.
- [3] Takemura, and S. Aoki, "Some characterizations of minimal Markov basis for sampling from discrete conditional distributions", Annals of the Institute of Statistical Mathematics, 2003.
- [4] Takemura, and S. Aoki,"Distance reducing Markov bases for sampling from discrete space", Bernoulli, to appear, 2005.
- [5] Semple, M. Steel, "Phylogenetics", volume 24 of Oxford Lecture Series in Mathematics and its Applications, Oxford University Press, Oxford, (2003).
- [6] Agnarsson, R. Greenlaw. "Graph theory, modeling, applications, and algorithms ", United States of America, (2007).
- [7] H. Abbass and H. S. Mohammed Hussein " On Toric Ideals for 3×n/3 - Contingency Tables with Fixed Two Dimensional Marginals, n is a multiple of 3", European Journal of Scientific Research, Vol. 123, pp. 83-98, 2014.
- [8] H. Abbass and H. S. Mohammed Hussein "Generating 3× n/3 -Contingency Tables Using The Action of Dihedral Group D\_n, n is Multiple of 6", J. of Karbala for Mathematics, (accept), 2014.
- [9] Felsenstein, "Inferring Phylogenies", Sinauer Associates, Sunderland, (2003).
- [10] M. Dyer, and C. Greenhill, "Polynomial-time counting and sampling of two-rowed contingency tables", Theoretical Computer Sciences, 246, (2000), pp. 265-278.
- [11] P. Diaconis, and B. Sturmfels, "Algebraic algorithms for sampling from conditional distributions", The Annals of Statistics, 26, pp. 363-397, (1998).
- [12] S. Aoki, A. Takemura, "The largest group of invariance for Markov bases and toric ideals", J. Symbolic Computation 43 (5) (2008) 342–358.
- [13] S. Aoki, and A.Takemura, "The list of indispensable moves of unique minimal Markov basis of  $3\times 4\times K$  and  $4\times 4\times K$  contingency tables with fixedtwo- dimensional

marginals", Technical Report METR 03-38, Department of Mathematical Engineering and Information Physics, The University of Tokyo. Submitted for publication, (2003).

- [14] S. Aoki, and A. Takemura, "Minimal basis for connected Markov chain over 3×3×K contingency tables with fixed two-dimensional marginals". Australian and New Zealand Journal of Statistics, 45, pp. 229-249, (2003).
- [15] S. Aoki, and A. Takemura," Minimal Invariant Markov Basis for Sampling Contingency Tables with Fixed Marginals", Annals of the Institut of Statistical Mathematics, pp.230-256, 2006.
- [16] S. Sullivant, J. Rauh, T.Kahle. "Postive Margins and Primary Decomposition", North Carolina State, pp.1-22, 2012.
- [17] W. J. Gilbert," Modern Algebra With Application ",John Wiley, Inc., New Jersey, U.S.A, 2004.

Volume 3 Issue 12, December 2014 www.ijsr.net