European Journal of Combinatorics 59 (2017) 169–186

Contents lists available at ScienceDirect

European Journal of Combinatorics

journal homepage: www.elsevier.com/locate/ejc

Finite phylogenetic complexity of $\mathbb{Z}_p$ and invariants for $\mathbb{Z}_3$ ☆

Mateusz Michałek

Polish Academy of Sciences, Warsaw, Poland

ARTICLE INFO

Article history:
Received 9 February 2016
Accepted 15 August 2016
Available online 7 September 2016

ABSTRACT

We study phylogenetic complexity of finite abelian groups—an invariant introduced by Sturmfels and Sullivant (2005). The invariant is hard to compute—so far it was only known for $\mathbb{Z}_2$, in which case it equals 2 (Sturmfels and Sullivant, 2005), (Chifman and Petrović, 2007). We prove that phylogenetic complexity of any group $\mathbb{Z}_p$, where p is prime, is finite. We also show, as conjectured by Sturmfels and Sullivant, that the phylogenetic complexity of $\mathbb{Z}_3$ equals 3.

1. Introduction

The motivation for our work comes from phylogenetics—a science that aims at reconstructing the history of evolution. We will not present here all the concepts from phylogenetics as they are not needed for the statement and the solution of the problem that we study. Let us just say that to any tree T and a finite abelian group G, by considering a Markov process on a tree, one associates a projective toric variety $X(T, G)$. The explicit description of the variety and the associated polytope is given in Definition 2.2. We refer the interested reader to [28,14,31,27,24], where the relations to phylogenetics and applications are explained in detail. The equations defining $X(T, G)$ are called phylogenetic invariants. In all the cases that we study, determining phylogenetic invariants for any tree T was reduced to so-called star or claw trees using toric fiber product [31, Theorem 26], [33, Corollary 2.11]. These trees, denoted by $K_{1,n}$ have one inner vertex and n leaves. Let us cite Draisma and Kuttler [12]:

"We have now reduced the ideals of our equivariant models to those for stars, and argued their relevance for statistical applications. The main missing ingredients for successful applications are equations for star models. These are very hard to come by (…)".

☆ Supported by a grant Iuventus Plus of the Polish Ministry of Science project 0301/IP3/2015/73.
E-mail address: wajcha2@poczta.onet.pl.

http://dx.doi.org/10.1016/j.ejc.2016.08.007
0195-6698/

In our previous work with Maria Donten-Bury [10] we have shown how to obtain phylogenetic invariants of bounded degree. However, it is highly nontrivial to obtain such a bound. To study these bounds Sturmfels and Sullivant defined two functions.

Definition 1.1 ($\psi(n, G)$, $\psi(G)$). Let $\psi(n, G)$ be the degree in which the (saturated) ideal defining $X(K_{1,n}, G)$ is generated. Let $\psi(G)$, called the phylogenetic complexity of G, be the supremum of $\psi(n, G)$ over $n \in \mathbb{N}$.

As observed by Sturmfels and Sullivant [31]: "The phylogenetic complexity $\psi(G)$ is an intrinsic invariant of the group G. (…) It would be interesting to study the group-theoretic meaning of this invariant". However, these invariants are very hard to compute. So far we only know $\phi(\mathbb{Z}_2) = 2$ [31,7]. Based on numerical computations Sturmfels and Sullivant proposed the following conjecture.

Conjecture 1.2 (*[31, Conjecture 29]*). *For any finite abelian group G we have $\psi(G) \leq |G|$.*

However, for $G \neq \mathbb{Z}_2$ we do not know if $\psi(G)$ is finite. Our first main theorem is as follows.

Theorem 1.3. *For any prime number p the phylogenetic complexity of $\mathbb{Z}_p$ is finite.*

Depending how general the model is there are other qualitative results on the degree of phylogenetic invariants. For very general, so-called equivariant models, the fact that on set-theoretic level there exists a bound was proved in [12,13,11]. For the class of G-models that includes all the models introduced in this article, on the level of projective schemes the bounds were obtained in [25]. Finally, for group-based models, but *only on Zariski open set*, the bound of the degrees by $|G|$ was proved in [6]. Our second main theorem is as follows.

Theorem 1.4. *The phylogenetic complexity of the group $\mathbb{Z}_3$ equals 3.*

This allows to find all phylogenetic invariants for any tree for the group $\mathbb{Z}_3$. As far as we know, this is the only model, different from the Jukes–Cantor model, where the complete list of phylogenetic invariants for any tree is obtained. For real data applications of phylogenetic invariants we refer for example to [29]. We would also like to mention that a related result was recently obtained by Donten-Bury in [9] on scheme-theoretic level.

The techniques that we use rely entirely on algebraic combinatorics. We present the above described problems in the combinatorial terms in Section 2. In different words, we study algebraic properties of a family of integral polytopes.

Although the original construction of varieties $X(T, G)$ was inspired by phylogenetics, recently they appeared in other sciences [19–21,32]. We would like also to mention that the varieties $X(T, G)$ share many other very interesting algebraic and combinatorial properties related to their Hilbert polynomial, normality and deformations [4,5,16,22].

The problems of the degrees in which toric ideals are generated appear in many different contexts [2]. Let us summarize the results and conjectures about group-based models in the following table.

	Group-based models				
Polynomials defining:	$\mathbb{Z}_2$	$\mathbb{Z}_3$	$\mathbb{Z}_2 \times \mathbb{Z}_2$	$\mathbb{Z}_p$	G
Gröbner basis	Degree 2 by [7]				Question 1.5
Generators of the ideal	Degree 2 by [31]	Degree 3 by Theorem 1.4	Conjecture [31, Conjecture 30]	Finite by Theorem 1.3	Conjecture 1.2 [31, Conjecture 29]
The projective scheme		Degree 3 [9]	Degree 4 [25]		Finite by [25]
Set-theoretically					Finite by [11]
On a Zariski open subset			Degree 4 [26]		Degree $\leq |G|$[6]