Afrika Matematika (2025) 36:138
https://doi.org/10.1007/s13370-025-01357-y

A Lie symmetry approach to migration and gene frequency

Sameerah Jamal[1,2]

Received: 11 April 2025 / Accepted: 23 July 2025

Abstract
This paper examines a genetic frequency model, governed by a second-order differential equation that includes a migration mean and variance parameters. Such a model is critical for understanding how gene flow affects genetic variation across migrating populations. We prove how to reduce the equation to a solvable expression, and thereafter apply Lie symmetries to obtain exact solutions. Finally, a case study of the CCR5-△32 mutant gene which provides resistance to HIV, is discussed.

Keywords Genetics · Symmetries · Initial conditions

1 Introduction

The study of population genetics is essential for understanding the mechanisms that drive the genetic makeup of populations. The processes of natural selection, genetic drift and mutation play major roles in shaping life. Kimura [8] is best known for his contributions to population genetic models, investigating the distribution and frequency of genetic variation within and between populations.

Such models describe the evolution of allele frequencies but are simplified for mathematical treatment. Most models are a continuous analogue of the classical branching process in genetics and is actually an approximation to it when the population is large. These models are in fact a subclass of the Kolmogorov equation or Fokker-Planck model [3, 15]. Genetics itself has numerous applications in the related sciences of biology, conservation, medicine and agriculture. Gene frequencies refer to the proportions of genes in a population [2]. There is randomness in reproduction leading to changes in gene frequencies, this randomness is often referred to as drift. Models aim to describe the dynamics of allele frequencies in

✉ Sameerah Jamal
sameerah.jamal@wits.ac.za

[1] School of Mathematics, University of the Witwatersrand, 2001, Johannesburg, South Africa

[2] DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), Johannesburg, South Africa

Published online: 20 August 2025

populations under the influence of genetic drift, selection, mutation, and other factors, using partial differential equations (PDEs) [17].

Consider the following migration model [8]

$$\frac{\partial u(x,t)}{\partial t} = \frac{a}{2}\frac{\partial^2}{\partial x^2}\{(x-m)^2\, u\,(x,t)\} + \beta\frac{\partial}{\partial x}\{(x-m)\, u\,(x,t)\}, \tag{1.1}$$

where $u(x, t)$ is the frequency distribution that the allele/gene frequency at time t (in generations) is x. The constants a and β describe that the migration rate $\bar{m}$ fluctuates randomly from generation to generation with mean β and variance a. We assume a diploid population consisting of a fixed number of individuals in each generation, and consider the alleles A_1 and A_2 each with respective frequencies x and $1 - x$. Hence, m is the frequency of allele A_1 in the migrants. Mathematically, the model describes how allele frequencies are distributed across the population at time t. Since x represents the allele frequency at a given point in the population, it can range from 0 to 1, where: $x = 0$ means the allele is absent in the population, while $x = 1$ means every individual in the population has this allele.

Lie symmetry studies are highly relevant because they provide a systematic framework for gaining insight into complex mathematical models that arise in many fields of science and engineering. Recent advances in this area exploit symmetries plus ansatz methods, for example, the analysis of the Hietarinta equation [16] and the time-dependent variable-coefficient Korteweg-de Vries equation [12]. Parabolic equations, which appear in many reaction-diffusion processes have been actively studied by symmetries, for example, [4, 5, 7]. In this paper, we showcase a powerful implementation of symmetries in genetics.

The structure of this work is organized as follows: Sect. 2 contains the relevant theory we require in this study. In Sect. 3, we present transformations that convert the genetic model into the one-dimensional parabolic heat equation, followed by the determination of exact solutions in Sect. 4. The approach used is the Lie invariant method, which incorporates the optimal system of one-dimensional subalgebras. Section 5 discusses the mathematical implications of our study, and we conclude in Sect. 6.

2 Preliminaries

A point symmetry is a transformation that maps all solutions of a given equation to another solution of the same equation. Consider the variables u^a, $a = 1, \ldots, q$ dependent on x^j, $j = 1, \ldots, p$. Suppose that

$$H_a(x, u^{(k)}) = 0, \tag{2.1}$$

is a system of differential equations, where $u^{(k)}$ represents the k^{th} derivative of u with respect to x. A one-parameter Lie group transformation with ε as the group parameter, invariant under equation (2.1), is defined as

$$\bar{x} = \Xi(x, u; \varepsilon) \quad \bar{u} = \Phi(x, u; \varepsilon). \tag{2.2}$$

Invariance of (2.1) under the transformation (2.2) dictates that any solution $u = \Theta(x)$ of (2.1) maps into another solution $v = \Psi(x;\varepsilon)$ of equation (2.1). The expanding of (2.2) around the identity $\varepsilon = 0$, provides the transformations:

$$\bar{x}^j = x^j + \varepsilon\xi^j(x,u) + O(\varepsilon^2), \tag{2.3}$$

$$\bar{u}^a = u^a + \varepsilon\eta^a(x,u) + O(\varepsilon^2). \tag{2.4}$$

The Lie group action is taken from the infinitesimal generators acting on the space of all variables. Let the symmetry generator be of the form

$$Y = \xi^j(x,u)\partial_{x^j} + \eta^a(x,u)\partial_{u^a}. \tag{2.5}$$

The invariance condition is

$$Y\left(H_a(x,u^{(k)})\right) = 0, \quad \text{when} \quad H_a(x,u^{(k)}) = 0, \tag{2.6}$$

where Y is must be prolonged on all derivatives of $H_a(x,u^{(k)})$. The operator in Eq. (2.5) defines the characteristics

$$\frac{dx^j}{\xi^j} = \frac{du^a}{\eta^a},$$

and the zero-order invariants

$$V^{[0]}(x^j,u^a). \tag{2.7}$$

3 A change of variables

We prove that transformations exist for rendering (1.1) to the classical heat equation. The purpose of changing variables is that we may exploit several unique features of the heat equation. The construction of such transformations originated by Sophus Lie [10] and we seek simple transformations that may be manipulated to solve the problem under study. Thus we establish the next result.

Theorem 1 *Equation (1.1) is converted to*

$$\omega_\tau - \omega_{yy} = 0, \tag{3.1}$$

which is the 1+1 heat equation, via the transformations

$$\tau = t, \tag{3.2}$$

$$y = \sqrt{2}\ln(x-m)\frac{1}{\sqrt{a}}, \tag{3.3}$$

followed by

$$u(y,\tau) = \omega(y,\tau)\mathrm{e}^{-\frac{\sqrt{2}(3a+2\beta)y}{4}\frac{1}{\sqrt{a}}-\frac{(a+2\beta)^2\tau}{8a}}. \tag{3.4}$$

Proof An application of (3.2) and (3.3), converts Eq. (1.1) to the form

$$\frac{\partial}{\partial t}u(y,\tau) - \frac{\partial^2}{\partial y^2}u(y,\tau) - \frac{\sqrt{2}(3a+2\beta)}{2}\frac{1}{\sqrt{a}}\frac{\partial}{\partial y}u(y,\tau) - (a+\beta)u(y,\tau) = 0. \tag{3.5}$$

Thereafter, the transformation (3.4) converts (3.5) to the heat equation (3.1). □

4 The solutions

We take two possible approaches to the solution of (1.1). Firstly, we solve using the optimal system of subalgebras for a description of nonequivalent classes of exact solutions. Secondly, we provide a formula to solve an initial value problem for the model. Both techniques are significant for the mathematical analysis of gene frequencies.

4.1 Classical invariant solutions

The Lie point symmetries of (3.1) are found by way of the procedure detailed in Section 2. They are:

$$\begin{aligned} X_1 = \frac{\partial}{\partial y}, \quad X_2 = \frac{\partial}{\partial \tau}, \quad X_3 = \omega\frac{\partial}{\partial \omega}, \\ X_4 = y\frac{\partial}{\partial y} + 2\tau\frac{\partial}{\partial \tau}, \quad X_5 = 2\tau\frac{\partial}{\partial y} - y\omega\frac{\partial}{\partial \omega}, \\ X_6 = 4\tau y\frac{\partial}{\partial y} + 4\tau^2\frac{\partial}{\partial \tau} - (y^2+2\tau)\omega\frac{\partial}{\partial \omega}, \\ X_\alpha = \alpha(y,\tau)\frac{\partial}{\partial \omega}, \end{aligned} \tag{4.1}$$

where $\alpha(y,\tau)$ is an arbitrary solution of (3.1).

The optimal one dimensional subalgebra for the above generators are [14]

$$X_1,\ X_2 + bX_3,\ X_2 - X_5,\ X_2 + X_6 + bX_3,\ X_4 + 2bX_3,$$

where $b \in \mathbb{R}$. Thus the exact solutions based on the above subalgebras are split into the five cases of the optimal system. A subalgebra yields invariant transformations which are substituted into (3.1) to find a reduced equation. The reduced equation may be solved and all transformations reversed to solve the original equation (1.1).

Solution (A) X_1

From this subalgebra we calculate the invariants $n = \tau,\ \omega = z\,(n)$, so that the exact solution of (3.1) is

$$\omega(y,\tau) = C_1, \tag{4.2}$$

where C_1 denotes an arbitrary constant henceforth. Theorem 1 allows us to reverse all transformations to find the exact solution for (1.1), viz.

$$u(x,t) = C_1\,\mathrm{e}^{-\frac{(3\,a+2\,\beta)\ln(x-m)}{2\,a} - \frac{(a+2\,\beta)^2 t}{8\,a}}. \tag{4.3}$$

Solution (B) $X_2 + bX_3$

In this case, the subalgebra yields the invariant functions $n = y,\ \omega = z\,(n)\,\mathrm{e}^{b\tau}$, and an invariant solution of (3.1) is

$$\omega(y,\tau) = \left(C_1\,\mathrm{e}^{\sqrt{b}y} + C_2\,\mathrm{e}^{-\sqrt{b}y}\right)\mathrm{e}^{b\tau}, \tag{4.4}$$

where C_2 is an arbitrary constant. From Theorem 1, we have the solution for (1.1), viz.

$$u(x,t) = \left(C_1\,\mathrm{e}^{\sqrt{2}\ln(x-m)\sqrt{b}\frac{1}{\sqrt{a}}} + C_2\,\mathrm{e}^{-\sqrt{2}\ln(x-m)\sqrt{b}\frac{1}{\sqrt{a}}}\right)\mathrm{e}^{bt}\mathrm{e}^{-\frac{(3\,a+2\,\beta)\ln(x-m)}{2\,a} - \frac{(a+2\,\beta)^2 t}{8\,a}}. \tag{4.5}$$

Solution (C) $X_2 - X_5$

The invariants here are $n = y + \tau^2, \omega = z\,(n)\,\mathrm{e}^{y\tau + \frac{2\,\tau^3}{3}}$, and once applied, the invariant solution of (3.1) is

$$\omega(y,\tau) = \left(C_1\,\mathrm{Ai}\left(\tau^2 + y\right) + C_2\,\mathrm{Bi}\left(\tau^2 + y\right)\right)\mathrm{e}^{y\tau + \frac{2\,\tau^3}{3}}, \tag{4.6}$$

where the functions *Ai*(*n*) and *Bi*(*n*) are the *Airy Ai* and *Airy Bi* functions, respectively. Hence by Theorem 1, the exact solution for (1.1) is.

$$u(x,t) = \left(C_1\,\mathrm{Ai}\left(\sqrt{2}\ln(x-m)\frac{1}{\sqrt{a}} + t^2\right) + C_2\,\mathrm{Bi}\left(\sqrt{2}\ln(x-m)\frac{1}{\sqrt{a}} + t^2\right)\right) \mathrm{e}^{t\sqrt{2}\ln(x-m)\frac{1}{\sqrt{a}} + \frac{2\,t^3}{3}}\,\mathrm{e}^{-\frac{(3\,a+2\,\beta)\ln(x-m)}{2\,a} - \frac{(a+2\,\beta)^2 t}{8\,a}}. \tag{4.7}$$

Solution (D) $X_2 + X_6 + bX_3$

The invariants of this case are found to be $n = \frac{4\,\tau^2+1}{4\,y^2}$, $\omega = z\,(n)\,\mathrm{e}^{\frac{b\left(4\,\tau^2+1\right)}{8\,\tau^2+2}\left(\left(-4\,\tau^2-1\right)c\arctan\left(\frac{1}{2\,\tau}\right)-2\,\tau y^2\right)} = \frac{1}{\sqrt{y}}$.

Thus, this subalgebra gives us the invariant solution of (3.1), that is

$$\omega(y,\tau) = \frac{\left(C_1\,M_{\frac{i}{4}b,\,\frac{1}{4}}\left(\frac{iy^2}{4\,\tau^2+1}\right) + C_2\,W_{\frac{i}{4}b,\,\frac{1}{4}}\left(\frac{iy^2}{4\,\tau^2+1}\right)\right)}{\sqrt{y}}\mathrm{e}^{\frac{1}{8\,\tau^2+2}\left(\left(-4\,\tau^2-1\right)c\arctan\left(\frac{1}{2\,\tau}\right)-2\,\tau y^2\right)}, \tag{4.8}$$

where $W_{\mu,\nu}(p)$ and $M_{\mu,\nu}(p)$ are Whittaker functions.

Thus, a reversal of the transformations from Theorem 1, yields the exact solution for (1.1), viz.

$$u(x,t) = \left(C_1 M_{\frac{i}{4}c,\,\frac{1}{4}} \left(\frac{2\,i\,(\ln(x-m))^2}{(4\,t^2+1)\,a} \right) + C_2\, W_{\frac{i}{4}c,\,\frac{1}{4}} \left(\frac{2\,i\,(\ln(x-m))^2}{(4\,t^2+1)\,a} \right) \right) \mathrm{e}^{2\,\frac{(\ln(x-m))^2}{a\left(8\,t^2+2\right)} \left(-1/2\,\frac{c\left(4\,t^2+1\right)a}{(\ln(x-m))^2} \arctan\left(1/2\,t^{-1}\right) - 2\,t \right)} \mathrm{e}^{-\frac{(3\,a+2\,\beta)\ln(x-m)}{2\,a} - \frac{(a+2\,\beta)^2 t}{8\,a}} \frac{1}{\sqrt{\sqrt{2}\ln(x-m)\,\frac{1}{\sqrt{a}}}}. \tag{4.9}$$

Solution (E) $X_4 + 2bX_3$

Here the invariants are $n = y\frac{1}{\sqrt{\tau}},\ \omega = z(n)\,\tau^b$. The symmetry invariants of the subalgebra produce the exact solution of (3.1) as

$$\omega(y,\tau) = \mathrm{e}^{-\frac{y^2}{4\,\tau}} \left(\frac{\tau}{y^2} \right)^{b-\frac{1}{2}} \left(U\left(b+1,\ \frac{3}{2},\ \frac{y^2}{4\,\tau} \right) C_2 + M\left(b+1,\ \frac{3}{2},\ \frac{y^2}{4\,\tau} \right) C_1 \right) y^{2\,b}, \tag{4.10}$$

where $U(\mu,\nu,p)$ and $M(\mu,\nu,p)$ are Kummer functions.

Finally, by Theorem 1, the exact solution for (1.1), is as follows

$$u(x,t) = \left(C_1 \sqrt{2}\ln(x-m)\,\mathrm{e}^{-\frac{(\ln(x-m))^2}{2\,at}} M\left(1+b,\ \frac{3}{2},\ \frac{(\ln(x-m))^2}{2\,at} \right) \frac{1}{\sqrt{a}} \frac{1}{\sqrt{t}} + C_2 \sqrt{2}\ln(x-m)\,\mathrm{e}^{-\frac{(\ln(x-m))^2}{2\,at}} U\left(1+b,\ \frac{3}{2},\ \frac{(\ln(x-m))^2}{2\,at} \right) \frac{1}{\sqrt{a}} \frac{1}{\sqrt{t}} \right) t^b \mathrm{e}^{-\frac{(3\,a+2\,\beta)\ln(x-m)}{2\,a} - \frac{(a+2\,\beta)^2 t}{8\,a}} \tag{4.11}$$

4.2 An initial condition solution

A second type of solution involves an initial condition. This initial condition is important because it helps define how the gene or allele frequency evolves from the beginning of time as it is influenced by migration factors. This solution requires the fundamental solution of the heat equation and its convolution [1, 6, 13]. Under Theorem 1, for the initial data problem $u(x,0) = g(x)$, coupled to (1.1), we find that the solution to (1.1) is given by the formula

$$u(x,t) = \left(\frac{1}{\sqrt{4\pi\tau}} \int_{\mathbb{R}} e^{-\frac{(y-m)^2}{4\tau}} g(m) e^{\frac{\sqrt{2}m(3a+2\beta)}{4\sqrt{a}}} dm \right) \mathrm{e}^{-\frac{\sqrt{2}(3\,a+2\,\beta)y}{4}\frac{1}{\sqrt{a}} - \frac{(a+2\,\beta)^2\tau}{8\,a}}. \tag{4.12}$$

For example, suppose that the initial condition is a fixed gene frequency, i.e. $u(x,0) = x_0$, then (4.12) yields the exact solution

$$u(x,t) = (x-m)^{-\frac{3a+2\beta}{2a}} \, \mathrm{e}^{\frac{(3a+2\beta)\ln(x-m)+2at(a+\beta)}{2a}} x_0. \tag{4.13}$$

A quick check for $t = 0$ gives the fixed gene frequency x_0 as expected, since the solution is subject to the inital condition.

5 A case study and parameter analysis

In Fig. 1, we explore an interesting graphical solution from the established solutions which describe the gene frequency versus probability density function for a given time. We assume that the frequency of A_1 in the migrants is 10%, i.e. $m = 0.1$ and that this is strictly less than the allele frequency. We also consider other typical values associated with migration.

Recall, that the mean migration rate β is the average proportion of individuals that are migrants—we use 15% as the mean. On the other hand, the variance a measures how much the migration rates across populations differ—we let the variance be 0.05.

It can be seen in the density functions in Fig. 1 that the highest density of the 2nd generation occurs when the gene frequency is close to 80%.

We now consider a specific population with alleles, A_1 and A_2, in a region where migration influences allele frequencies. This case study will examine how migration affects allele frequencies over time, specifically focusing on changes due to gene flow from a neighbouring population.

In a real-world study [11] of the CCR5-△32 allele which provides resistance to HIV, in different populations across Europe affected by historical migration patterns, found that the CCR5-△32 allele is more frequent in Northern Europe compared to Southern Europe, likely caused by Viking era migrations.

The chemokine receptor CCR5 is a coreceptor for the macrophage-tropic strains of HIV-1 [11]. Let the allele A_1 be a mutant allele of the CCR5 gene named △32 linked to a strong resistance against infection by HIV, and correspondingly, A_2 is a wild-type allele.

Consider Table 1 with population samples from Northern and Southern Europe. Let N be the effective size of the population, and assume a variance value for allele A_1 of 2. Individuals were noninfected and unrelated, and allele frequencies are extracted from [11].

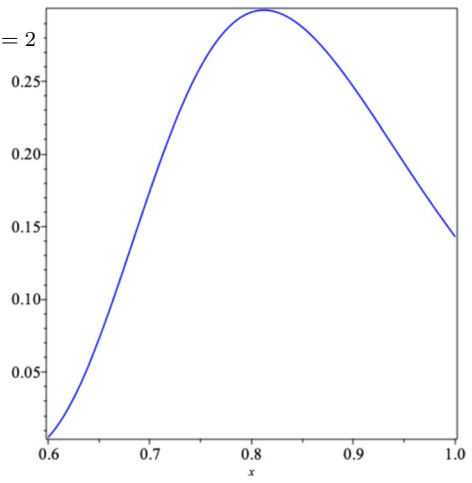

Fig. 1 Solution (C) for $C_1 = 1, C_2 = 0, m = 0.1, a = 0.025, \beta = 0.15, t = 2$

Table 1 Data set of allelic frequencies at different European locations

Locus	N	CCR5-△32 frequency	Wild-Type frequency	Estimated mean migration rate
Northern Europe	743	0.13	0.87	0.05
Southern Europe	6308	0.08	0.92	0.02

Table 2 CCR5-△32 frequency density predictions in Southern Europe

Locus	t=1	t=10	t=50	t=100
Southern Europe	0.09	0.17	2.65	87.73

This scenario may be ideally captured by the model under study, which is capable of predicting allelic frequency distribution changes through evolving generations. As a demonstration, consider data from Table 1 and say, the solution (4.12). That is, suppose that the population starts at the A_1 gene frequencies of Table 1.

We calculate the frequency densities predicted by the model and construct Table 2. It is assumed that Nordic populations in the North disseminated the A_1 gene [9].

From Table 2, we see that the alleles increasingly coexist in the Southern Europe population, over time. By the 100th generation, the frequency density of the allele is expected to increase 1000-fold. The results confirm that migration introduces individuals carrying different allele combinations, altering the allele frequencies in the Southern population. This is a hopeful result in driving geographic resistance to HIV.

6 Conclusion

This paper presents an investigation of the transformations for a population genetics model. Initially, a coordinate transformation was used to reduce the model to the heat transfer model, followed by a second transformation chosen to provide similarity solutions. Ultimately, we demonstrated that this approach leads to solutions for the model, and that frequency distributions may be obtained and interpreted. A real-world data set of a a mutant allele of the CCR5 gene was adopted to display the predictive power of the results. In future work, we shall pursue other factors that affect population dynamics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s13370-025-01357-y.

Acknowledgements Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the CoE-MaSS.

Funding Open access funding provided by University of the Witwatersrand. None.

Data Availability Not Applicable.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Use of AI tools declaration The author declares that they have not used artificial intelligence tools in the creation of this article.

References

1. Evans, L.C.: Partial Differential Equations. Graduate Studies in Mathematics, vol. 19, 2nd edn. American Mathematical Society, Rhode Island (2010)
2. Feller, W.: Diffusion Processes in Genetics. In: Neymaln, J. (ed.) Second Berkeley Symposium on Mathematical Statistics and Probability, pp. 227–247. University of California Press, Berkeley (1951)
3. Fokker, A. D.: Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld. Annalen der Physik **348** (1914)
4. Jamal, S.: Prototype gliomas: invariant solutions and expected survival times. Disc. Contin. Dyn. Syst. - S **18**(4), 882–893 (2025)
5. Jamal, S.: Solving an isotropic grey matter tumour model via a heat transfer equation. Open Phys. **23**, 20250120 (2025)
6. Jamal, S., Maphanga, R.: A progressive approach to solving a CEV model via symmetry invariant surface conditions. AIMS Math. **9**(2), 4326–4336 (2024)
7. Jamal, S., Champala, R., Khan, S.: Lie symmetries and the invariant solutions of the fractional Black–Scholes equation under time-dependent parameters. Fract. Fract. **8**, 269 (2024)
8. Kimura, M.: Diffusion models an population genetics. J. Appl. Probab. **1**, 177–232 (1964)
9. Libert, F., et al.: The CCR5 mutation conferring protection against HIV-1 in Caucasian populations has a single and recent origin in Northeastern Europe. Hum. Mol. Genet. **7**, 399 (1998)
10. Lie, S.: On integration of a class of linear partial differential equations by means of definite integrals. Arch. Math. Naturvidensk **5**, 328–358 (1881). (**(German)**)
11. Lucotte, G.: Distribution of the CCR5 gene 32-basepair deletion in West Europe. A hypothesis about the possible dispersion of the mutation by the vikings in historical times. Hum. Immunol. **62**(9), 933–936 (2001)
12. Mabenga, C., Muatjetjeja, B., Motsumi, T.G.: Multiple soliton solutions and other solutions of a variable-coefficient Korteweg–de Vries equation. Int. J. Mod. Phys. B **37**(09), 2350090 (2023)
13. Maphanga, R., Jamal, S.: A terminal condition in linear bond-pricing under symmetry invariance. J. Nonlinear Math. Phys. **30**, 1295–1304 (2023)
14. Olver, P.J.: Applications of Lie Groups to Differential Equations, 2nd edn. Springer, New York (1993)
15. Planck, M.: Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie. Sitzungsber Preuss Akad. Wiss. Phys. Math. **K1**, 325 (1917)
16. Seele, M.F., Muatjetjeja, B., Motsumi, T.G., Adem, A.R.: A novel study of the (2+1)-dimensional Hietarinta equation with a trivial balancing number. Int. J. Theor. Phys. **64**, 98 (2025)
17. Wright, S.: Evolution in Mendelian populations. Genetics **16**, 97–159 (1931)