

# **Rethinking the “gypsy” retrotransposon: A roadmap for community-driven reconsideration of problematic gene names**

Kevin H-C Wei<sup>1\*</sup>, Reema Aldaimani<sup>1</sup>, Dat Mai<sup>1</sup>, Daniel Zinshteyn<sup>2</sup>, PRV Satyaki<sup>3</sup>, Justin Blumenstiel<sup>4</sup>, Erin S Kelleher<sup>5</sup>, Ethel C Brooks<sup>6\*</sup>†

<sup>1</sup> Department of Integrative Biology, University of California Berkeley, Berkeley, CA

<sup>2</sup> Molecular Therapeutics Program, Fox Chase Cancer Center, Philadelphia, PA

<sup>3</sup> Whitehead Institute for Biomedical Research, Massachusetts Institute of Technology, Cambridge, MA

<sup>4</sup> Department of Biology and Biochemistry, University of Houston, Houston, TX

<sup>5</sup> Department of Ecology and Evolution Biology, University of Kansas, Lawrence, KS

<sup>6</sup> Department of Women’s, Gender, and Sexuality Studies, Rutgers University, New Brunswick, NJ

\*Corresponding authors

†Person of Romani heritage

## **Abstract**

Transposable elements (TE) can replicate and mobilize at the detriment of the host genome they reside in; thus, they are considered selfish or parasitic genetic elements. First discovered in *Drosophila*, the gypsy retroelement was named in reference to the migratory history of the Romani people. This name later came to denote an entire family of wide-spread TEs and related viruses. Here, we discuss why the continued use of “gypsy” in this scientific context is insensitive and perpetuates racial stereotypes. Further, we outline a series of steps for the reconsideration of problematic gene names that take into account the relevant scientific communities, literature continuity, and, importantly, the harmed communities.

## **What are transposable elements and why are they selfish?**

Transposable elements (TEs), also known as mobile genetic elements, are genes that can move around in the genome. Typically, they encode proteins that can cut themselves out from one location or make copies of themselves and insert into different locations. The latter are known as retrotransposons, as their replication requires the reverse transcription of RNA back into DNA prior to insertion. TEs were discovered by the venerable Barbara McClintock in the 40s and 50s when studying mosaic purple pigmentation in maize [1]. She uncovered genetic elements that can move and in doing so affect the expression of gene controlling purple pigment production. At the time, McClintock's discovery was largely dismissed [2]. The profundity of her findings was not fully appreciated until decades later, when researchers began realizing how prevalent these mobile elements are. Four decades after her discovery, Dr. McClintock won the Nobel Prize in Physiology or Medicine for this research.

Since the discovery by McClintock, decades of research have revealed TEs to be a major and abundant constituent of the genomes of most eukaryotic phyla [3]. Half of the human genome, for example, is composed of TEs [4]. Their existence and ubiquity are partly due to their unique mode of propagation. Unlike normal genes which are passed down from parents to offspring through reproduction, TEs can, in addition, directly make copies of themselves. Therefore, they can perpetuate and successfully replicate in genomes even if they do not contribute to organismal function. This has garnered them the reputation of being "selfish genetic elements" [5] and "genomic parasites" [6].

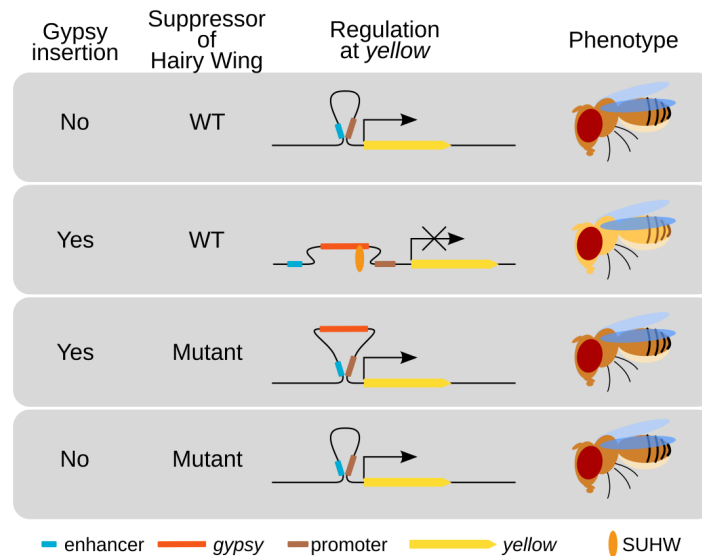
However, the self-replicative mechanism is a source of genetic mutations and inherently problematic to the organism. When TEs transpose, they cause DNA breakage, and when landed into genes, they can disrupt gene function leading to genetic abnormalities and diseases. To date, there are more than 120 genetic diseases that result from TE insertions [7]. In cases where TE insertions are unmitigated, organisms often become completely sterile due to overwhelming mutagenic assault [8]. Therefore, sophisticated genomic defense mechanisms

are in place to keep TEs inactive and the strict control of TEs is essential to maintaining genome stability.

### **What is the “gypsy” retrotransposon?**

In 1983, the same year that McClintock received her Nobel prize, Modolell, Bender, and Meselson reported the discovery of a novel TE as the underlying basis for the puzzling genetic phenomena known as suppressible mutations [9] (Box). They found that a TE caused different mutant phenotypes in fruit flies (*Drosophila*) by inserting around different genes. Because of the mobile nature of this element, they called the TE “gypsy” in reference to the commonly held stereotype that gypsy people move frequently. Coincidentally, Gerasimova *et al*, a group of scientists from the former Soviet Union, concurrently discovered the same TE and named it *mdg4*, for mobile dispersed gene 4 [10]. As is often the case when there are multiple names for the same gene, the more evocative name, “gypsy,” was adopted by the genetics community. Since then, TEs with sequence similarity to *gypsy* have been widely identified in plants and vertebrates, making this group of TEs one of the most ubiquitous types of transposable elements. This large group of elements, known as a superfamily, is collectively known as *gypsy* or *gypsy-like* after the founding element in flies [3].

Soon after the discovery of “gypsy,” it was revealed that this TE behaves much like a RNA virus and replicates with the RNA-mediated copy and paste mechanism, making it a retrotransposon [11,12]. Further evolutionary analyses revealed that many RNA viruses are highly related to other *gypsy-like* retrotransposons and may have originated from one of the members of this family [13,14]. Thus, the *gypsy* superfamily encompasses not only a diverse and prevalent set of retrotransposons but also a family of viruses (*Metaviridae*) [15,16].



### Box: The genetic puzzle of suppressible mutations [9]:

Fruit flies, or *Drosophila*, are popular among geneticists not only for how easy they are to manipulate and raise in the lab, but also the wealth of genetic mutants with visible deformities that researchers have collected and studied since the early 1900s. Some of these mutations cause flies to have lighter body color, malformed wings, or fewer bristles, typically due to inactivation of the genes that are needed for normal development of these traits. A puzzling group of mutants are known as suppressible mutants: the malformation associated with these mutants can be reversed in the presence of another mutation. Modolell, Bender, and Meselson discovered that, in many cases, the genes with suppressible mutations have the same novel sequence insertion nearby. They isolated this sequence and named it “*gypsy*”.

When a *gypsy* element inserts near a gene, it can block the interaction between regulatory elements necessary for the gene’s activity. This “insulating effect” is the result of a protein (encoded by the gene Suppressor of Hairy wing or SuHW) binding to the *gypsy* sequence. In the absence of this protein, the *gypsy* insertion no longer insulates the regulatory activity, allowing the gene to be properly expressed thus suppressing the mutant phenotype.

### Is “gypsy” a pejorative?

The word “gypsy” is considered a pejorative for people of Romani heritage. It carries many negative connotations such as swindling or untrustworthy. These negative connotations are evident from the colloquial use of “gypped” which means to be cheated or tricked. It’s important to recognize and put into perspective that the Roma are both historically and currently marginalized. Surveys by the European Union Agency for Fundamental Rights reports that 80%

of Roma in Europe are at risk of poverty [17]. To many people of Roma heritage, the word “gypsy” is a racial slur that evokes an ugly history of dehumanization, persecution, enslavement, and genocide. And this word continues to be used for their subjugation and oppression [18]. Many advocacy groups such as Amnesty International [19] and the European Roma Rights Centre [20] have denounced the use of this word. The casual use of this word perpetuates many of the discriminatory stereotypes put onto the Roma people.

### **Why is using “gypsy” as the name of transposable elements harmful?**

*“It’s a name that many of us didn’t choose for ourselves, one that has been used against us to vilify and demean our characters and culture/ethnicity without any second impression because people will only take it at name value. To have a gene [reflect] all of our worst stereotypes (selfish, detrimental to others around it, etc) doesn’t sit right with me and probably a lot of others.”*

*S.L., person of Romani heritage.*

The discoverers of the “gypsy” element named it after the traveling nature of the element and had no derogatory intent (Personal communications with WB). For similar reasons, many TEs are given names that express movement like *roamer*, *vagabond*, *voyager*, *pogo*, and *wanderer*. However, “gypsy” is uniquely problematic. First, because TEs are commonly characterized as parasitic, selfish, and deleterious, this coinage perpetuates the discriminatory stereotypes associated with the word and the people. Second, because repression of TEs are pivotal to genome stability and organismal fitness, this usage further alludes to offensive ideologies that the people need to be controlled or sequestered. Third, the stereotype that “gypsies” are migratory is itself a product of historical persecution, as they were forced to keep moving [21].

Lastly, despite being the largest minority group in Europe, people of Roma heritage are currently severely underrepresented in STEM and higher education; in Europe only 1% of Roma enter higher education [22]. Therefore, the continued usage of this word, especially in a

scientific context, can act as an unnecessary deterrent and barrier to entry. This is particularly dire as TEs, as major constituents of genomes, are commonly taught in introductory Genetics classrooms.

### **TEs can also be beneficial. Why are we only focusing on their negative attributes?**

This is indeed the case. There are many examples of TEs generating beneficial change for the host organism. One textbook example is the melanization of the peppered moth during the industrial revolution in Britain. More melanized individuals became better adapted due to improved camouflage on tree barks that were darkened by pollution; these dark variants are caused by a TE inserted into the gene producing dark pigmentation. Undoubtedly, just like other types of mutations, TEs are sources of genetic diversity and innovations.

However, splitting hairs on the benefits and detriments of the *gypsy* element and other TEs is missing the bigger picture. Even if these elements were all beneficial, the usage of this pejorative as a name of a TE invokes stereotypes of the Roma people that have historically led to persecutions and continue to denigrate to this day.

### **My Romani friends aren't offended by the term "gypsy", so why should we be?**

We cannot speak for all Romani people so there is a possibility that some are unbothered by or even proud of this term. While we do not wish to invalidate the view of these Romani people, we also cannot ignore those who do find its usage in this context disparaging. Changing the gene name will do no harm to the Romani people who find no issue with this usage, while at the same time, would begin to reduce the harm it may have caused and validate the perspectives of the people offended by it.

### **This probably doesn't affect many people. Does it even matter?**

Perpetuated stereotypes and racism have insidious and long-lasting effects. This is true even if they do not cause "direct" harm. Given how severely under-represented the Romani people are in science and higher education, it is likely that very few in the Roma community are exposed to

this issue. However, for any person of Romani heritage who aspires to enter science, this word and its associations can act as a deterrent. For this barrier to put off even one person is to do a gross disservice to the core value that science is universal and equitable.

### **How are gene names typically changed?**

Gene names are typically given by the person to first discover them. The field of *Drosophila* Genetics has long cherished the whimsical tradition of researchers selecting evocative and tongue-in-cheek names for new genes. This has led to adored names like *hedgehog* and *ZELDA*. However, there are also examples of poorly-considered choices. The gene, *fruitless*, required for male flies to recognize female flies was initially named “*fruity*” [23]. This was because male flies with mutations in the gene court other males. However, this name, a derogatory slang for gay individuals, was subsequently recognized as poorly considered and corrected in a series of publications by Jeffrey C. Hall [24,25].

Similar issues are also plaguing other scientific communities; entomologists and ornithologists [26,27] are also grappling with insensitive species names [28]. In parallel to our arguments, the Entomological Society of America has recently announced the retirement of the common name of *Lymantria dispar* - the “gypsy” moth [29] - a species of moth invasive to North America, and well known for its environmental and economic damage to hardwood forests of Eastern US and Canada. To dispel the unfortunate association to the Roma people, the society has officially changed the common name to the spongy moth. Similarly, regional ornithological societies have been evaluating names of birds that are considered to be culturally insensitive [30]. Outside of human genetics which established a committee that approves gene names and symbols (the HUGO Gene Nomenclature Committee), the greater genetics community has no formal governing body for gene names. The precedent set by *fruitless* illustrates that name changes can be grassroots - driven by individuals within the community.

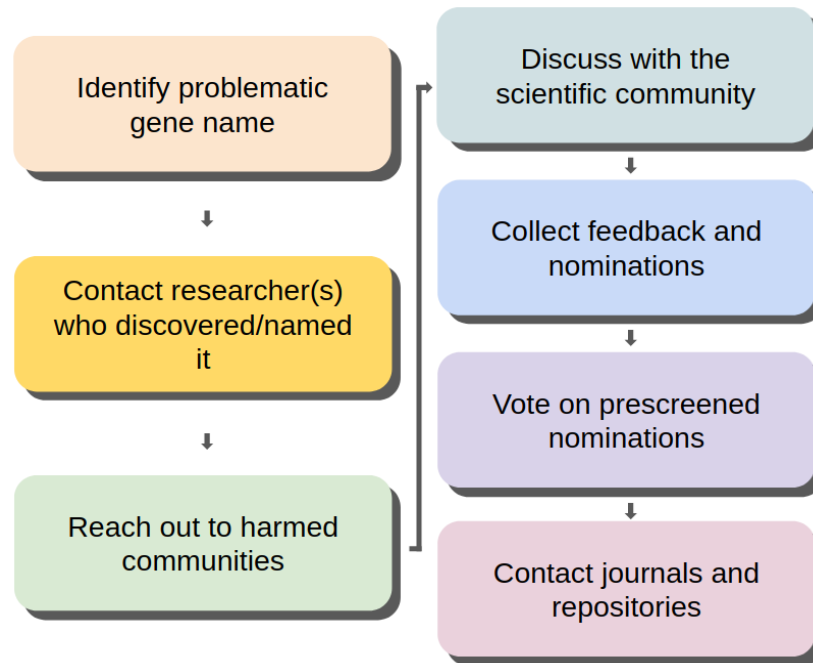


The usage of “gypsy” for some TE types has far extended beyond *Drosophila* genetics as it, rather than just one gene, has come to represent an entire family of TE, as well as viruses, found all across the tree of life. Beyond the study of TEs, “gypsy” elements have other interesting molecular properties that are being studied as important regulatory mechanisms in genomes. Therefore, to change this name, a community-driven push is necessary to effect wide-spread adoption and lasting change. A concerted, open, and well-documented effort will also minimize the initial confusion that can come with the transition and ensure continuity in the scientific literature.

### **How do we move forward?**

Because of the scope and history of this issue, we have taken a series of steps that pay respect to the originator of the coinage, consider the experience and reaction from Roma communities, and weigh the feedback from the scientific community. Balancing all these considerations, we believe that a nomination followed by voting will put the onus not on a single person or group of people to be the arbiter, but on the community to collectively come to an acceptable and appropriate replacement. Once a new name is elected by the community-led process, journals, repositories, and editors can then be contacted to reinforce this change. And as reviewers, we can collectively recommend the use of the new name. At the same time, we also recommend the usage of Ty3 [31] as the family name and reinstatement of *mdg4* as the name of the *Drosophila* TE, given the history of the co-discovery [10].

As the genetics community navigates through cultural sensitivities, political climates, and human experiences, there will likely be increased scrutiny on unfortunate products of legacy practices. Although the steps we are taking here to raise awareness and change “gypsy” may not be fully applicable to all cases, they provide a well-considered roadmap for future endeavors.



**Figure 1. Roadmap for reconsidering and renaming of offensive gene names.**

### **Acknowledgements**

We sincerely thank Welcome Bender, Irina Arkhipova, Daniel Barbash, Harmit Malik, and Doris Bachtrog for insightful discussions; Mariana Wolfner, Cedric Feschotte, and Leigh Greenwood for comments on the manuscript; Alexander Suh and the 2021 Uppsala TE Symposium for allowing us to present on this issue; the community on the TE World Slack Channel for lively discussions and suggestions; and the community and mods on the subreddit r/romani for their kind support and critical feedback.

## References

1. McClintock B. The Origin and Behavior of Mutable Loci in Maize. *Proc Natl Acad Sci U S A*. 1950;36: 344–355.
2. Comfort NC. “The real point is control”: the reception of Barbara McClintock’s controlling elements. *J Hist Biol*. 1999;32: 133–162.
3. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhou B, et al. A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*. 2007;8: 973–982.
4. Deniz Ö, Frost JM, Branco MR. Regulation of transposable elements by DNA modifications. *Nat Rev Genet*. 2019;20: 417–431.
5. Doolittle WF, Sapienza C. Selfish genes, the phenotype paradigm and genome evolution. *Nature*. 1980;284: 601–603.
6. Orgel LE, Crick FH. Selfish DNA: the ultimate parasite. *Nature*. 1980;284: 604–607.
7. Hancks DC, Kazazian HH Jr. Roles for retrotransposon insertions in human disease. *Mob DNA*. 2016;7: 9.
8. Kidwell MG, Kidwell JF, Sved JA. Hybrid Dysgenesis in *DROSOPHILA MELANOGASTER*: A Syndrome of Aberrant Traits Including Mutation, Sterility and Male Recombination. *Genetics*. 1977;86: 813–833.
9. Modolell J, Bender W, Meselson M. *Drosophila melanogaster* mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3-kilobase mobile element. *Proc Natl Acad Sci U S A*. 1983;80: 1678–1682.
10. Gerasimova TI, Ilyin YV, Mizrohki LJ, Semenova LV, Gerogiev GP. [MOBILITY OF THE MOBILE DISPERSED MDG4 GENETIC ELEMENT AND ITS ROLE IN THE APPEARANCE OF UNSTABLE MUTATIONS IN *DROSOPHILA*]. *Proc Natl Acad Sci U S S R*. 1983;271: 977–980.
11. Marlor RL, Parkhurst SM, Corces VG. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. *Mol Cell Biol*. 1986;6: 1129–1134.
12. Kim A, Terzian C, Santamaria P, Péliesson A, Purd’homme N, Bucheton A. Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. *Proc Natl Acad Sci U S A*. 1994;91: 1285–1289.
13. Xiong Y, Eickbush TH. Origin and evolution of retroelements based upon their reverse transcriptase sequences. *EMBO J*. 1990;9: 3353–3362.
14. Malik HS, Eickbush TH. Modular evolution of the integrase domain in the Ty3/Gypsy class of LTR retrotransposons. *J Virol*. 1999;73: 5186–5190.
15. Pringle CR. Virus taxonomy--1999. The universal system of virus taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses

during 1998. Arch Virol. 1999;144: 421–429.

16. Hull R. Classification of reverse transcribing elements: a discussion document. Arch Virol. 1999;144: 209–13; discussion 213–4.
17. Second European Union Minorities and Discrimination Survey Roma – Selected findings. In: European Union Agency for Fundamental Rights [Internet]. 10 Nov 2016 [cited 3 Feb 2022]. Available: <https://fra.europa.eu/en/publication/2016/second-european-union-minorities-and-discrimination-survey-roma-selected-findings>
18. Reference Paper – antigypsyism.eu. [cited 3 Feb 2022]. Available: [https://antigypsyism.eu/?page\\_id=17](https://antigypsyism.eu/?page_id=17)
19. The Roma in Europe: 11 things you always wanted to know but were afraid to ask. In: Amnesty International [Internet]. 23 Apr 2015 [cited 3 Feb 2022]. Available: <https://www.amnesty.org/en/latest/campaigns/2015/04/roma-in-europe-11-things-you-always-wanted-to-know-but-were-afraid-to-ask/>
20. Terminology. In: European Roma Rights Centre [Internet]. [cited 3 Feb 2022]. Available: <http://www.errc.org/what-we-do/advocacy-research/terminology>
21. Machemer T. These Moths Will Be Renamed to Stop Use of an Ethnic Slur. Smithsonian Magazine. 12 Jul 2021. Available: <https://www.smithsonianmag.com/smart-news/these-moths-will-be-renamed-stop-use-ethnic-slur-180978151/>. Accessed 19 May 2022.
22. Higher Education Internationalisation and Mobility : Research projects : Centre for Higher Education and Equity Research (CHEER) : University of Sussex. [cited 3 Feb 2022]. Available: <http://www.sussex.ac.uk/education/cheer/researchprojects/rise>
23. Gill KS. A mutation causing abnormal courtship and mating behavior in males of *Drosophila melanogaster*. Am Zool. 1963;3: 507.
24. Hall JC. Courtship among males due to a male-sterile mutation in *Drosophila melanogaster*. Behav Genet. 1978;8: 125–141.
25. Gailey DA, Hall JC. Behavior and cytogenetics of fruitless in *Drosophila melanogaster*: different courtship defects caused by separate, closely linked lesions. Genetics. 1989;121: 773–785.
26. Campbell V. Toward Inclusivity in Birding: Forum Discusses Renaming Eponymous Birds. In: All About Birds [Internet]. 23 Apr 2021 [cited 3 Feb 2022]. Available: <https://www.allaboutbirds.org/news/toward-inclusivity-in-birding-forum-discusses-renaming-eponymous-birds/>
27. AOS. FAQ: AOS Community Congress on English Bird Names. In: American Ornithological Society [Internet]. 14 Apr 2021 [cited 3 Feb 2022]. Available: <https://americanornithology.org/english-bird-names/faq-aos-community-congress-on-english-bird-names/>
28. Lancette J. Breaking Barriers in Entomology: The Better Common Names Project. Am

Entomol. 2021;67: 10–11.

29. Better Common Names Project. [cited 3 Feb 2022]. Available: <https://www.entsoc.org/publications/common-names/better-common-names-project>
30. Driver RJ, Bond AL. Towards redressing inaccurate, offensive and inappropriate common bird names. *Ibis* . 2021;163: 1492–1499.
31. Clark DJ, Bilanchone VW, Haywood LJ, Dildine SL, Sandmeyer SB. A yeast sigma composite element, TY3, has properties of a retrotransposon. *J Biol Chem*. 1988;263: 1413–1423.