Image: Nathalie Vessillier

# Gephebase:
# Connecting traits to genes
# in plants and animals

www.gephebase.org

# Gephebase: Connecting traits to genes in plants and animals

Mutations form the raw bulk of heritable variation upon which traits evolve. What are the genes and the mutations that underlie the evolution of drug tolerance in malaria parasites, or of insecticide resistance in their mosquito hosts? Which genes make an agricultural crop suitable for a given soil or climate? How do species adapt to anthropogenic challenges, such as climate change, or polluted environments? Which gene should be modified to obtain a particular character of interest in a plant species? Those are examples of lingering questions in the field of evolutionary genetics, with both fundamental and societal implications, which rely on our understanding of the link between the traits observed in living organisms (the phenotype) and the genes. Each month, dozens of publications report the elucidation of the actual genes and mutations responsible for measurable differences between individuals and species. Gephebase is a universal, single resource integrating our accumulated knowledge of more than 1700 genetic loci of natural, domesticated and experimental phenotypic variation. Each entry corresponds to a genetic change associated to a phenotypic difference, together with bibliographical information, molecular details and taxonomic information. The data is available at **www.gephebase.org** and is easily searchable using Boolean operators on the different fields, and exportable in spreadsheet format.

**It is important to gather data from multiple species into one database**

A "Canis" Gephebase search results in 37 entries, with known genes linked to size, pigmentation, skeletal traits, hair features, and physiology. Data volume could likely triple in the next five years. While a dog-specific database could compile these data in order to make dog genetics increasingly predictive, it would miss the opportunity to compare with other organisms as Gephebase does. For instance, both dogs and humans adapted to starch-rich diets in the Neolithic by multiplying the number of copies of a salivary enzyme gene (see "amylase" entries in Gephebase). Similar parallels exist between determinants of size (IGF1 gene) and pigmentation (Agouti gene) with humans and other animals.

**Predicting the genetic basis of xenobiotic resistance**

An evolutionary experiment of unprecedented scale is taking place on the planet, as organisms are quickly evolving resistance to xenobiotics of human origin such as drugs, insecticides, herbicides, and environmental pollutants. It is now both of practical and fundamental importance to catalog these cases and assess the predictability of their genetic basis. Our dataset is the only existing resource compiling the cases of evolution of resistance among parasites (eg. infectious yeasts, malaria), vectors (eg. mosquito) and agricultural pests (eg. blights, insects, weeds,

**Fig. 1.** *Ear of wheat. During the history of domestication several crops have been selected for larger grain size. Mutations in the same gene (GW2) have been selected independently in wheat and rice to increase grain size. The GW2 gene thus appears to be a good target for genetic manipulation in other crops to increase grain size.*
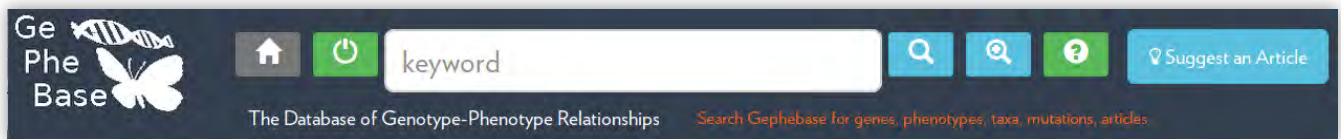
**Fig. 2.** *Gephebase. an online database of the relationships between genes and visible characteristics in animals and plants. www.gephebase.org*

nematodes) to the chemical arsenal that humans have deployed against them. Compilation of current data reveals that independent mutations in the same gene is very often responsible for the evolution of resistance to the same toxic compound in various species. To test whether a particular pollutant has been released in a given environment, it can thus be more practical and more efficient to screen living organisms for the mutations conferring resistance rather than to search for the pollutant itself.

**Identifying the DNA sites that should be tweaked to produce a trait of interest**
It is often said that most of the world food consists of three cereals – wheat, corn, and rice (178 Gephebase entries). Importantly, the genetic basis of desirable traits in one crop can usually guide research in another species, highlighting the importance of Gephebase as a comparative tool not limited to a single species. Our database includes a rich array of entries related to other agronomic plants such as tomato, sunflower, soybean, potato, and cultivated fruits. Preliminary data on domesticated plants reveal a prevalence of parallel evolution, with independent mutations in the same gene underlying evolution of the same traits in various crops (eg. blood oranges independently involved the selection of a transposon insertion in the promoter of the same gene, in Italy and China). The prevalence of such cases of parallel evolution highlights how gephebase can help to predict the genes that should be mutated to produce a trait of interest. The new CRISPR-Cas9 genetic engineering technology allows researchers to rapidly modify the DNA sequence of domestic plants and animals as they wish. Gephebase thus appears as a

great resource to identify the sites within the DNA that should be modified to produce traits of interest.

Our database is currently fed on a voluntary basis by a few curators. We are now looking for solutions and opportunities to maintain and expand the database, to spark community-wide interest, and to foster meta-analyses of the current gephebase data to clarify the dictionary of correspondence between genes and traits.

**www.gephebase.org**

Further reading

1. Martin A, Orgogozo V (2013) The loci of repeated evolution: a catalogue of genetic hotspots of phenotypic variation. Evolution 67(5):1235–1250. PDF

2. Orgogozo, V, Morizot, B, & Martin, A (2015). The differential view of genotype–phenotype relationships. Frontiers in Genetics, 6, 179. PDF

3. Orgogozo V, Peluffo A, Morizot B. (2016) The "Mendelian gene" and the "molecular gene": two relevant concepts of genetic units. Current Topics in Developmental Biology 119:1-26. CTDB PDF

4. Martin A, Courtier-Orgogozo V, Morphological evolution repeatedly caused by mutations in signaling ligand genes. Diversity and Evolution of Butterfly Wing Patterns : An Integrative Approach. Eds. Sekimura T, Nijout F. Springer. Book chapter PDF
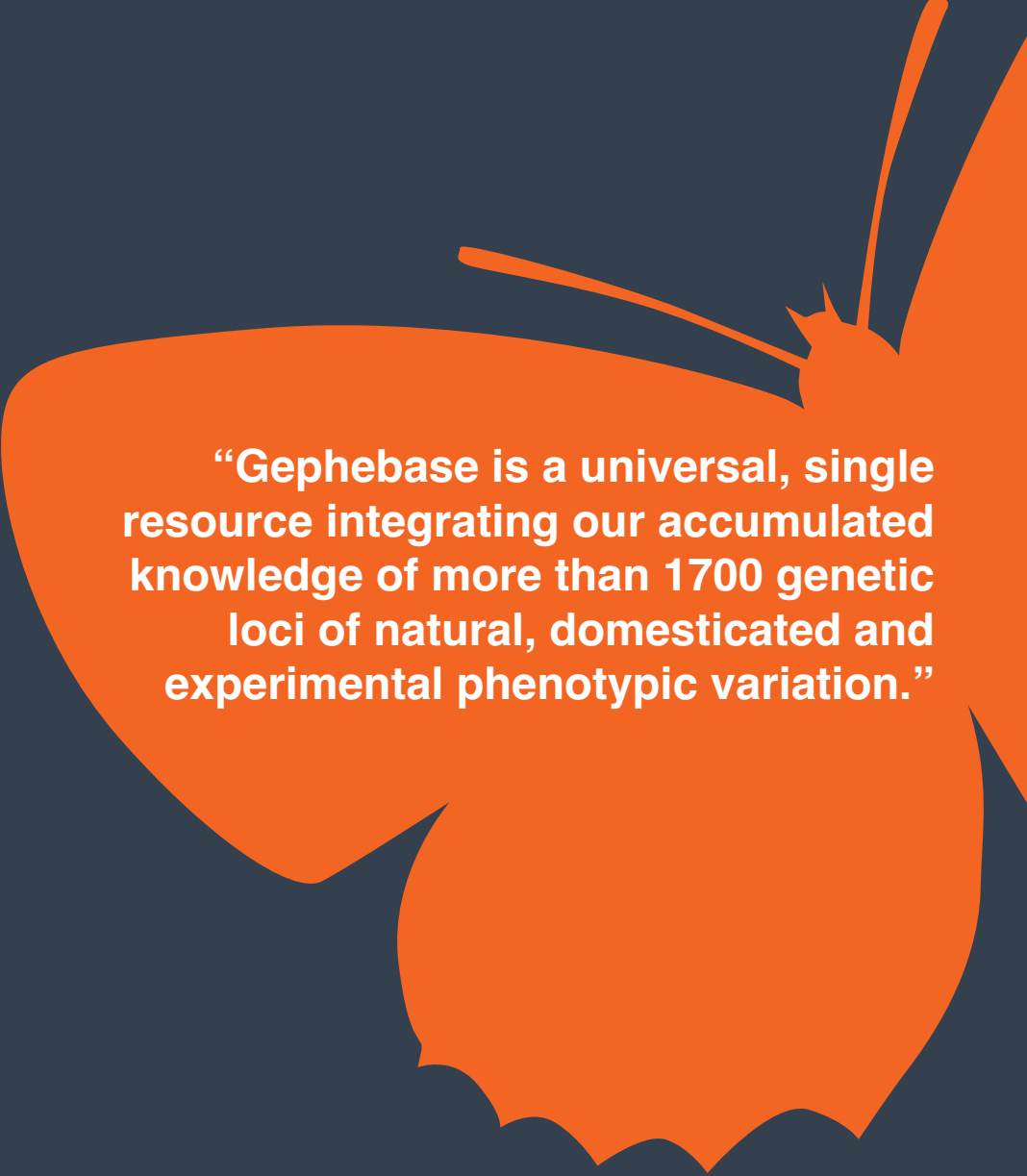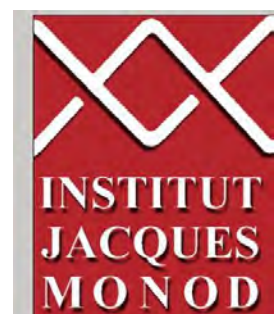
**Fig.3.** *Overview of the "Advanced search" page of Gephebase. One example of search is shown here: a search for 10- to 99-bp deletions in cis-regulatory regions that have been associated with changes in morphological traits in Mammals.*

Virginie Courtier-Orgogozo
Directeur de recherche CNRS
Institut Jacques Monod
Université Paris Diderot
Tel: +33 1 57 27 80 43
virginiecourtier.wordpress.com
virginie.courtier@ijm.fr

Arnaud Martin
Assistant Professor
Georges Washington University
Tel: +1 202 994 2384
https://dnacrobatics.com/
arnaud@gwu.edu

"Gephebase is a universal, single resource integrating our accumulated knowledge of more than 1700 genetic loci of natural, domesticated and experimental phenotypic variation."