SIMPLIFICATION OF MENDELIAN FORMULAE

PROFESSOR W. E. CASTLE

BUSSEY INSTITUTION, HARVARD UNIVERSITY

Castle, W. E. 1913. Simplification of Mendelian formulae. *The American Naturalist*, 47:170-182.

Electronic Scholarly Publishing Project

Foundations Series: Classical Genetics

Series Editor: Robert J. Robbins

The ESP Foundations of Classical Genetics project has received support from the ELSI component of the United States Department of Energy Human Genome Project. ESP also welcomes help from volunteers and collaborators, who recommend works for publication, provide access to original materials, and assist with technical and production work. If you are interested in volunteering, or are otherwise interested in the project, contact the series editor: rrobbins@fhcrc.org.

Bibliographical Note

This ESP edition, first electronically published in 2002, is a newly typeset, unabridged version, based on the original 1913 edition published in the *The American Naturalist.*

Production Credits

Scanning of originals: ESP staff
OCRing of originals: ESP staff
Typesetting: ESP staff
Proofreading/Copyediting: ESP staff
Graphics work: ESP staff
Copyfitting/Final production: ESP staff

Castle, W. E. 1913. Simplification of Mendelian formulae. *The American Naturalist,* 47:170-182.

SIMPLIFICATION OF MENDELIAN FORMULAE

PROFESSOR W. E. CASTLE

BUSSEY INSTITUTION, HARVARD UNIVERSITY

Professor Bessey in his recent presidential address* expresses the opinion that Mendelian terminology is needlessly complicated. This opinion most biologists will heartily endorse, and not a few Mendelians will be among their number. For those who work most extensively with Mendelian formulae feel most keenly the need of simplification in these the tools of their investigations.

Professor Morgan, in the January NATURALIST,† makes a commendable effort to introduce reforms. I desire heartily to endorse his effort, but would suggest certain modifications in method.

The Mendelian may say in justification of existing usage that it has arisen naturally step by step as knowledge of Mendelian phenomena has advanced, but this is of course no justification of its continued use, if it has become a hindrance rather than a help in the further advance of knowledge.

Morgan clearly points out the two historical steps by which present usage was reached. The first of these was Mendel's original recognition of segregating dominant and recessive characters existing in contrasted pairs, and his convenient designation of the former by capitals and of the latter by small letters. This usage answered perfectly so long as only a single modification of any character came under consideration, and indeed Mendel's observations did not go beyond this. But this system broke down when characters more complex in nature came under observation, as for example when Cuénot showed that more than a single differential factor exists between gray mice and albino mice. (2)

* *Science,* January 3, 1913.

† Vol. 47, pp. 5-16.

The ingenious and useful "presence and absence" hypothesis of Bateson was the second step which led to our present usage. On this hypotheses gray in mice is not the allelomorph of white, but of *no-gray*; while the allelomorph of white is color, or more properly speaking white is equivalent to *no-color* and *this* is the allelomorph of color.

Both of these steps have been amply justified by their utility in making possible the prediction of the previously unpredictable consequences of particular crosses.

It was natural that in applying the presence and absence hypothesis the usage of Mendel should have been retained, in accordance with which capital letters were used as the symbols of dominant characters and small letters as the symbols of recessive characters. But this retention has involved most unfortunate consequences and is, I believe, the real seat of our present difficulty.

Mendel's small letters stood for realities as truly as did the capitals. His *A* was a round form of pea, his *a* was a wrinkled form of pea; his *B* was a yellow-seeded, his *b* a green-seeded pea. But the significance of these terms has been changed under the presence and absence hypothesis. *A* still means a round pea, but *a* is simply a *not-round* pea; it may or may not be wrinkled. Likewise *B* is still a yellow-seeded pea, but *b* is nothing but a *not-yellow* pea; it may or may not be green under the presence and absence hypothesis. For all that *b* signifies now, the pea may be blue, violet, indigo or carmine.

It is most unfortunate, therefore, that the small letters, having lost their original significance, were not discarded altogether, for under the presence and absence hypothesis they have done nothing but cause mischief.

The investigator who employs them starts out well intentioned, with a clear notion that the small letters stand for negation only, that they are merely signboards to show what characters he is talking about, but presently, unless he is unusually careful, we find him talking about them as if they stood for *something*, instead of nothing; he speaks of repulsions and couplings or associations between *a* and *B*, or even between *a* and *b*. Think of it! How can *something* be coupled with *nothing?* How can *nothing* be inseparably bound up with *nothing?* It seems to me the consequent effect on inheritance is *absolutely "nothing"!*

Not only do the small letters thus lead to confusion of thought, they also tend to make formulae needlessly cumbersome, for they call for the use of *two* symbols for every character difference dealt with. These two symbols also are so much alike that both printer and reader are in momentary danger of confusing them, with the consequence that what is *is not*, and what is not *is*!

The small letters are not indispensable to accurate and exhaustive analysis of Mendelian phenomena, or to lucid exposition of them. See, for example, the fundamental researches of Cuénot into the color inheritance of mice, and his classic "notes" describing them. Like Cuénot, I have not found the use of the small letters necessary; but among nearly all other Mendelians the double terminology has become so nearly universal that a different usage seems almost to demand an apology. Indeed Lang* has suggested that such offenders against uniformity as Cuénot and I should be haled before an International Congress and be *directed* to conform; since which time I had almost abandoned hope of ever seeing improvement in the current confusing system, but Morgan's protest and proposal gives me new courage.

What we need first of all to symplify [*sic*]our present usage is to *abandon the dual terminology*. Where we are dealing with a *single* set of variations, let a *single* set of symbols suffice. Let us give up either the small letters or the large ones, it matters not which. If we retain *A*, then we have no need of *a*, for it is not, as Morgan at one time seems to assert and at another to deny, the "residuum" when *A* is lost; it means on the presence and absence hypothesis nothing but this, that *A* is not present. The rest of the organism is the "residuum." Morgan points out and his paper illustrates amply how under the dual system "the letters used may unintentionally come to stand for different things." The obvious thing to do, if we attempt reform, is to omit the superfluous symbol, either the small letter or the large one.

Morgan, however, clings to the dual nomenclature, but suggests a reversal of the usual significance. Thus the factor for pink-eye, he assumes, is present only in animals which are *not* pink-eyed, and the factor for black body color, he suggests, is present in all sorts of flies *except* those which are black bodied. This is confusion worse confounded.

But, seriously, I do not see that it is possible to improve the existing terminology, so long as we use two terms of opposite significance with reference to a single germinal variation. Certainly merely reversing the significance of existing terms will not do it. What we need first of all is *one set of symbols*, used in a single significance.

If this reduction is allowed, then I think that another aspect of Morgan's proposition might be extremely useful, viz., *that a mutation which behaves as a recessive in crosses be designated by a small letter.* This proposition was put into effect more than three years ago in a paper dealing with color inheritance in mice, though Morgan does not

* *Zeitsch. f. ind. Abstammungs- und Vererbungslehre*, 4, p. 40, 1910.

seem to have observed it. See Castle and Little (1909).* In the paper cited, three recessive color factors of mice were designated by small letters, viz., "*d*, the dilution factor"; "*s*, the factor which causes *spotting* with white;" and "*p*, the pink-eye (or paucity) factor."

In that same paper all *dominant* color factors of mice were designated by capitals. This seems to me a very necessary complement to the use of small letters to express recessive variations, and is in entire harmony with Mendel's original usage. But neither of these proposals can help matters much, *unless we discard the duplicate set of symbols, which is the chief cause of present confusion*. Thus if we use *s* for spotting, then we have no occasion to use *S* for no-spotting. We simply leave out all reference to spotting, and we shall understand that there is none, but that the normal condition prevails.

To be very explicit, my proposals for simplification of Mendelian terminology are three:

1. To abolish the current dual terminology and use only *one symbol*, where a single variation from the normal is involved.
2. To use a *small letter* to designate the factor responsible for a variation which is *recessive* in crosses with the normal.
3. To use a *capital letter* to designate the factor responsible for a variation which is *dominant* in crosses with the normal.

These proposals were made in substance in publications of the year 1909 and are here renewed under encouragement of Morgan's suggestive paper. Let us see how they would work if applied to the cases enumerated by Morgan. The eye color series described by Morgan, *l. c.*, page 13, involving three recessive mutations, is as follows:

	Revised Terminology	Morgan's Terminology
Red	*normal*	*PVE*
Vermilion	*v*	*PvE*
Pink	*p*	*pVE*
Pink-vermilion	*pv*	*pvE*
Eosin	*e*	*PVe*
Vermilion-eosin	*ve*	*Pve*
Pink-eosin	*pe*	*pVe*
Pink-vermilion-eosin	*pve*	*pve*

* *Science*, N. S., Vol. 30, pp. 312-314.

The revised terminology is obviously shorter and simpler. It is obtained by merely omitting the capital letters from Morgan's terminology, letters which stand only for negations. The symbols used are suggestive of the names employed for the various color categories of eyes, whereas in Morgan's terminology the most conspicuous symbols are suggestive only of other categories than the true one.

The revised terminology is more convenient than Morgan's in calculating the expected result of any mating, and it is equally reliable. The result of every possible mating within the series can be readily computed without the confusing presence of the large letters.

To those who have grown accustomed to the presence and absence terminology the objection will suggest itself that in naming the recessive character and ignoring its allelomorph, we are naming an absence or negative and disregarding what is present and positive. But this does not follow. Because a character is recessive it does not follow that it is negative. I quite agree with Morgan that the physiological condition which produces an eosin eye is as real as that which produces a vermilion, a pink or a red eye, and no mere negation; it is simply different. It is quite impossible to decide, from its behavior as a dominant or recessive in crosses, whether a character is positive or negative. This I have pointed out elsewhere (1911) and the same view has been repeatedly expressed by Shull. We have on record many instances in which one and the same character may behave at one time as a dominant, at another time as a recessive.

Our terminology may well recognize the dominant or recessive behavior of a variation, without implying anything as to its positive or negative nature, which must in many cases be conjectural or possibly non-existent. Different gradations of color, such as we have in the eye-series of Drosophila described by Morgan, may result merely from quantitative variations in cell constituents and consequent activities, nothing being lost. This idea concerning the possible nature of Mendelian factors in general I have developed elsewhere, concluding that "it is the substantial integrity of a quantitative variation from cell-generation to cell-generation that constitutes the basis of Mendelism. All else is imaginary."*

Morgan applies his altered system of nomenclature also to the body-color series and wing mutation series which he has discovered. This nomenclature we may simplify, as we did in the case of the eye-color series, without impairing its utility.

* *American Naturalist*, Vol. 46, p. 358, June, 1912.

Body-Color Series

	Revised Terminology	Morgan's Terminology
Wild fly …….	*normal*	*YBES*[6]
Yellow …….	*y*	*yBES*
Yellow-black …….	*yb*	*ybES*
Ebony …….	*e*	*YBeS*
Sable …….	*s*	*YBEs*[7]

Wing-Mutation Series

	Revised Terminology	Morgan's Terminology
Wild fly …….	*normal*	*MR*
Miniature …….	*m*	*mR*
Rudimentary …….	*r*	*Mr*
Rudimentary-miniature …	*mr*	*mr*

The taste of the reader will govern his choice between these two systems. Doubtless either can be used successfully, though the revised terminology seems to me preferable on the ground of simplicity and suggestiveness.

In the series with which Morgan has dealt, all the mutations under consideration are *recessive* in character, so that one can read the names of the varieties directly from his formulae, if one disregards altogether his large letters and pays attention only to the small ones. To insure this I have suggested omitting the large ones.

But if one were to extend Morgan's terminology to a series in which *dominant* mutations as well as recessive ones occur, hopeless confusion would result. For here some of the large letters would stand for mutations, while others would stand for the negation of mutations, so that without a key constantly at hand the formulae would be unusable.

If, however, we use the *single* system of symbols as I have suggested, a series which includes both dominant and recessive mutations, may be handled without confusion. In this case every symbol is significant, and its dominant or recessive character is indicated by the symbol, whether large or small. For example, consider the mouse-color series as described by Castle and Little (1909). In the

* For simplicity I here use *E* instead of Morgan's *Eb*.

† Morgan's list here contains *S*, but I suspect this is a misprint for *s*; if so, it is a living witness to the dangers of the dual system

paper cited, nine color factors were described, three of which clearly recessive have already been mentioned, viz., *d*, *p* and *s*. The remaining six were considered dominant factors. Mr. Little has since suggested, and I think with good reason, that one of them had better be omitted, since its existence has not been demonstrated beyond question. The six as given were *C*, the color factor; *Y*, the yellow factor; *Br*, the brown factor; *B*, the black factor; *R*, the restriction factor (producing a yellow coat); and *A*, the agouti or gray factor.

Mr. Little would omit either *C* or *Y*, since it has not been shown beyond question that the effects which had previously been ascribed to these two are not due to one and the same agency.

With the eight symbols which would remain, three being small letters, the others being or beginning with capitals, it is possible to write, without duplication of terms, formulae descriptive of the entire color series. But in so doing it would be necessary to designate the original or wild form in terms of factors supposed to be lost in its derivatives, and which have only come to light through such loss. This, as Morgan points out, involves redescribing the wild form every time a mutation arises and should be avoided if possible. I therefore favor Morgan's suggestion that each mutation as it arises be given some suitable descriptive name, the initial or other significant letter of which shall be its symbol. If, as is commonly true, the mutation is recessive in crosses with the wild or original type, its symbol will be a small letter. But if the mutation is dominant,* its symbol should be a large letter.

The original or wild type *need not be described in terms of its mutations*, as every duplicate system of terminology, even Morgan's, requires. The system would accordingly be capable of indefinite expansion without constant remodeling.

I favor Morgan's further suggestion that as new forms arise through recombination of simple "mutations" these be described, so far as possible, in terms of the simple mutations composing them. This principle is clearly illustrated in the names chosen by Morgan for the eye color series of Drosophila. It is surprising how little change this system necessitates in the common names with which we are already familiar, for example, in the mouse-color series.

The color mutations† of mice with which I am personally familiar number seven. If all of these are independent, *i.e.*, not "coupled" or "associated," there should be theoretically possible 127 different

* I have met two dominant mutations in guinea-pigs, one in rabbits, and one in mice, so that they can scarcely be called rare.

† I use the term mutation in the sense of unit-factor variation, not in that of DeVries.

combinations involving one or more of them. A considerable proportion of these combinations has been produced in my laboratory in the course of the last twelve years, the earlier and simpler ones by Dr. G. M. Allen or myself, the later and more complex ones by Mr. Little, who has in press an extensive paper dealing with his investigations. I shall deal with the series as known up to 1909. The historical order of appearance of the mutations is now unknown; I shall place them in the alphabetical order of the symbols used. It is also unknown whether each of them arose directly from the wild type. More probably they did not, but experiment shows that they might have done so, since each behaves in crosses as if it had a distinct and independent basis in the germ-plasm.

Wild Type and its Seven Mutations

1. *Wild* = gray.
2. *a* = albino (transmitting gray in crosses).
3. *b* = black.
4. *c* = cinnamon.
5. *d* = dilute.
6. *p* = pink-eyed.
7. *s* = spotted.
8. *Y* = yellow.

The *a* mutation, however combined, if present in a homozygous condition, prevents the development of pigment in the skin, hair or eyes. The *d* mutation, under like circumstances, makes the pigmentation of the coat dilute, or pale; the *p* mutation reduces even more strongly the pigmentation of coat and eyes alike, but does it in a different way; the *s* mutation causes pigment to be altogether wanting in certain areas of the coat more or less definite in position and extent, which areas accordingly appear as white spots.

Combinations of these four mutations present no difficulties of description or recognition, though breeding tests alone suffice to differentiate the several sorts of albinos, since all look alike. The nomenclature also is perfectly simple. Thus,

ap = albino transmitting the pink-eye mutation in crosses.

adp = albino transmitting both dilution and pink-eye in crosses, etc.

Combinations of *b*, *c*, and *Y*, one with another, form the fundamental and best known color varieties, which will now be considered.

In the *b* mutation, the fur is black; in the *c* mutation, it is brownish gray, called cinnamon. In the *Y* mutation, the coat is yellow. Of the several mutations mentioned, *Y* alone is dominant over the wild gray, but it occurs only in a heterozygous state, and hence never breeds true.

The complete color series involving these three mutations, but excluding all others, is as follows:

Wild = gray.

b = black.

c = cinnamon.

bc = black-cinnamon (chocolate).

Y = yellow (giving also gray offspring).

bY = black-yellow (giving also black offspring).

cY = cinnamon-yellow (giving also cinnamon offspring)

bcy = black-cinnamon-yellow (giving also black-cinnamon off spring).

To express the modification which this series undergoes if the *d* mutation is added to it, we need only prefix the symbol *d* to each of the formulae given and omit the term wild as no longer applicable. The series then becomes

d = dilute gray.

db = dilute black.

dc = dilute cinnamon.

etc.

Similarly an added *p* factor gives us the series

p = pink-eyed gray.

pb = pink-eyed black.

pc = pink-eyed cinnamon.

etc.

Also an added *s* gives us the series

s = spotted gray.

sb = spotted black.

sc = spotted cinnamon.

etc.

Adding both *d* and *p* gives us the series

dp = dilute pink-eyed gray.

dpb = dilute pink-eyed black.

dpc = dilute pink-eyed cinnamon.

etc.

Adding *d* and *s* gives us the series

ds = dilute spotted gray.

dsb = dilute spotted black.

dsc = dilute spotted cinnamon.

etc.

Adding *p* and *s* gives us the series

ps = pink-eyed spotted gray.

psb = pink-eyed spotted black.

psc = pink-eyed spotted ‘cinnamon.

etc.

Adding simultaneously *d*, *p* and *s*, gives us the series

dps = dilute pink-eyed spotted gray.

dpsb = dilute pink-eyed spotted black.

dpsc = dilute pink-eyed spotted cinnamon.

etc.

We thus secure eight different variations of the fundamental color series, or a total of sixty-four colored varieties. By prefixing *a* to the formula for each of these varieties, we obtain formulae for sixty-four different types of albinos, which though all looking alike (being snow white), yet would transmit in crosses the characteristics each of a different one of the sixty-four colored varieties.

We have thus accounted for the entire one hundred and twenty-eight variations which theoretically should result from recombining seven distinct mutations with the original form from which they sprang, and this has been done in relatively simple terms. Only one formula in the whole 128 contains as many as seven letters. *This is adpsbcY,* and would be read “an albino transmitting dilute pink-eyed spotted chocolate and dilute pink-eyed spotted yellow.” All the other formulae would contain from one to six letters. The current presence and absence system would require *sixteen letters* in every one of the 128 formulae to express the same facts, and the same letter would in some of the formulae be a capital and in others a small letter, so that the constant close attention of the reader would be required to decide in each case whether a particular mutation was or was not present. Morgan’s system would be only slightly less cumbersome for it would require in each formula fourteen instead of sixteen letters, and the same confusion would result from the presence of duplicate large

and small letters. The mere statement of these facts is sufficient to show that Mendelians can easily simplify their formulae and make themselves more readily intelligible to each other and to their fellow biologists, if they are only willing to do so.

There is another reason why I favor Morgan's terminology (as here simplified); it commits us to no physiological theory, but simply states facts. We are not required to suppose that the wild form contains a number of factors which by mutation have been lost. We may still do so, but we are not *forced* to do so. We are free to suppose with Morgan that merely a "readjustment" has taken place, and to make no assumption as to its nature, unless we choose to do so. This course does not prejudice the investigator of the physiology of color production but leaves him free to frame such hypotheses as will from his point of view best meet the situation. He is not bound down, for example, to a hypothesis of chromogen and ferments and so tempted with Riddle to throw over all Mendelism simply because Mendelians have in his opinion misinterpreted chemical facts.

That terminology evidently is most desirable which states demonstrated facts most clearly and simply, and makes fewest assumptions as to their explanation. Otherwise the investigator may be led to conclusions based on his terminology rather than his facts, and this can lead only to disaster.